DOCUMENT RESUME

ED 322 932 IR 053 333

| | |
|---|---|
| TITLE | Library Programs. Evaluating Federally Funded Public Library Programs. |
| INSTITUTION | Office of Educational Research and Improvement (ED), Washington, DC. |
| REPORT NO | LP-90-755 |
| PUB DATE | Jul 90 |
| NOTE | 112p. |
| AVAILABLE FROM | Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402. |
| PUB TYPE | Reports - Evaluative/Feasibility (142) |
| EDRS PRICE | MF01/PC05 Plus Postage. |
| DESCRIPTORS | *Evaluation Methods; Federal Aid; *Federal Programs; Institutional Evaluation; Library Administration; *Library Planning; Library Services; Library Standards; *Measurement Techniques; Policy Formation; Program Proposals; *Public Libraries |
| IDENTIFIERS | Library Services and Construction Act; Output Measures for Public Libraries |

ABSTRACT

Following an introduction by Betty J. Turock, nine reports examine key issues in library evaluation: (1) "Output Measures and the Evaluation Process" (Nancy A. Van House) describes measurement as a concept to be understood in the larger context of planning and evaluation; (2) "Adapting Output Measures to Program Evaluation" (Douglas L. Zweizig) presents 12 measures related to commonly occurring library objectives; (3) "Beyond Output Measures" (George D'Elia) proposes extending the evaluation process to include analysis of outcomes; (4) "Improving State Library Evaluation of Federal Programs" (Charles R. McClure) recommends ways for state library agencies to improve their evaluations; (5) "Evaluating the Impact of Federally Funded Public Library Youth Programs" (Mary K. Chelton) explores variables that hinder evaluation efforts; (6) "Qualitative and Quantitative Evaluation: Eight Models for Assessment" (Turock) identifies strengths and weaknesses of various evaluation methods; (7) "The Potential Role of Public Library Accreditation for Evaluating Federally Funded Library Programs" (Leigh Estabrook) provides background information about adapting accreditation to the public library; (8) "FSCS and the Evaluation of LSCA" (Mary Jo Lynch) defines current and possible future roles of the Federal-State Cooperative System; and (9) "The National Diffusion Network: Its Potential for Libraries" (Ellen Altman and Philip M. Clark) recommends ways of improving the evaluation designs of federally funded projects. (SD)

***********************************************************************
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
***********************************************************************

IR

# Library Programs

---

# Evaluating Federally Funded Public Library Programs

**Betty J. Turock**
**Project Director**

**July 1990**

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402

**U.S. Department of Education**
Lauro F. Cavazos
*Secretary*

**Office of Educational Research and Improvement**
Christopher T. Cross
*Assistant Secretary*

**Library Programs**
Anne J. Mathews
*Director*

**Information Services**
Sharon K. Horn
*Director*

July 1990

4

## FOREWORD

Anne J. Mathews

The Education Summit in Charlottesville, Virginia, in September of 1989 focused the attention of the President and the Nation's Governors on America's education system. There is concern about how well the system is serving its citizens, and a great deal of interest in making it more effective. As partners in the education endeavor, it seems prudent for librarians to renew their interest in examining their own programs and services. Together with the State Library Agencies, the Office of Library Programs, as the major federal funding source for libraries, shares this interest and seeks to promote improved practices in evaluation.

The following papers were commissioned to examine some of the key issues in library evaluation. The topics cover a wide spectrum of concerns ranging from the impact of the Federal-State Cooperative System for Public Library Data Collection (FSCS) to how accreditation can assist the evaluation process. Further the papers examine various aspects of the existing structures of evaluation, identify needs, and explore possibilities for meeting those needs.

Representing diverse points-of-view; some papers will generate more discussion than others. However, all are thought-provoking. Each author takes a unique stand. Each author presents important information. From this wealth of ideas and information, we hope to see new approaches, new methods, and new structures developed. We want--and we need--better ways of evaluating library programs and services.

It is our expectation that the ideas presented in the papers that follow will move us forward in improving the evaluation of library programs and services. As public attention turns our way, we must be prepared to demonstrate what libraries are doing and doing well.

On behalf of the Office of Library Programs, I would like to thank Betty J. Turock of Rutgers University, and Christina Dunn, Senior Associate in the Office of Library Programs, for the time and energy they put into bringing these papers together. They worked closely with the eminent researchers who wrote the papers, as well as with the staff of the Office of Library Programs, to bring the project to fruition.

## ACKNOWLEDGEMENTS

During the 1988 - 1989 academic year, while on leave from Rutgers University's School of Communication, Information and Library Studies, Betty J. Turock worked on the issues surrounding the evaluation of federally funded library programs not only with the staff of the Office of Library Programs (LP) in the U.S. Department of Education; but also with the Chief Officers of State Library Agencies (COSLA); with the Coordinators of the Library Services and Construction Act (LSCA) for the 50 state libraries; and with their Discussion Group in the Association of State and Cooperative Library Agencies (ASCLA), a division of the American Library Association (ALA). Their interest in evaluation was made obvious by their attendance at the series of conferences and meetings held during the year. Their combined assistance in the preparation of the White Papers was invaluable.

An advisory council of distinguished librarians served as a Focus Panel offering input at each stage of the work. Members were: Richard Cheski, State Librarian of Ohio; Blane Dessy, Director of Alabama's Public Library Service; Ray Ewick, Director of the Indiana State Library; June Garcia, Deputy Director of the Phoenix (AZ) Public Library and representative of the Public Library Association; Edwin Gleaves, State Librarian and Archivist of Tennessee; Wayne Johnson, State Librarian of Wyoming; Bridget Lamont, Director of the Illinois State Library; James Nelson, State Librarian and Commissioner of the Kentucky Department of Libraries and Archives; Larry Nix, Director of the Bureau of Library Development at the Wisconsin Department of Public Instruction; Sharon Rothenberger, Director of the Library Development Division at the Michigan State Library; Gary Strong, State Librarian of California; Barbara Weaver, Assistant Commissioner and State Librarian in New Jersey's Department of Education; and Nancy Zussy, State Librarian at the Washington State Library.

Robert Klassen, Director of Public Library Support at LP, the final member of the council, became an ongoing advisor. He scheduled frequent briefings with his staff, Adrienne Chute, Clarence Fogelstrom, Donald Fork, Dorothy Kittel, Evaline Neff, Sandy Pemberton, and Trish Skaptason, to keep the work directed toward productive channels.

The cogent comments of the peer reviewers were produced within a short timeframe. In addition to Adrienne Chute and Christina Dunn from LP and Sharon Rothenberger of the Michigan State Library, they were: Judith M. Foust, Deputy Director of the Brooklyn Public Library; Norman Horrocks, Vice-President of Scarecrow Press; Edwin S. Holmgren, Director of the New York Public Library; Jane Robbins, Director of the School of Library and Information Studies at the University of Wisconsin-Madison; and Kay Vandergrift, Associate Professor, Rutgers University School of Communication, Information and Library Studies.

Particular thanks are extended to Christina Dunn, who reviewed, edited, and prepared the final manuscript, and Zondra Carroll who prepared and produced the camera art for printing. Without their expert skills, these papers might never have been published.

# CONTENTS

| | | |
|---|---|---|
| Foreword | Anne J. Mathews | iii |
| Acknowledgements | | iv |
| Introduction | Betty J. Turock | 1 |
| Output Measures and the Evaluation Process | Nancy A. Van House | 5 |
| Adapting Output Measures to Program Evaluation | Douglas L. Zweizig | 15 |
| Beyond Output Measures | George D'Elia | 27 |
| Improving State Library Evaluation of Federal Programs | Charles R. McClure | 39 |
| Evaluating the Impact of Federally Funded Public Library Youth Programs | Mary K. Chelton | 55 |
| Qualitative and Quantitative Evaluation: Eight Models for Assessment | Betty J. Turock | 67 |
| The Potential Role of Public Library Accreditation for Evaluating Federally Funded Library Programs | Leigh Estabrook | 83 |
| FSCS and the Evaluation of LSCA | Mary Jo Lynch | 91 |
| The National Diffusion Network: Its Potential for Libraries | Ellen Altman and Philip M. Clark | 99 |
| About the Contributors | | 109 |

# INTRODUCTION

**Betty J. Turock**

Throughout 1988, the Public Library Support Staff from the U.S. Department of Education, Office of Educational Research and Improvement (OERI), Library Programs (LP) traveled about the country conducting regional workshops to improve the quality of technical assistance supplied from state library agencies to local community public libraries. During the sessions Program Officers were repeatedly petitioned by Library Services and Construction Act (LSCA) Coordinators and other staff members from the fifty state libraries for help in improving the evaluation of federally funded library programs. At the same time, in frequent informal dialogues as well as in formal semi-annual meetings and reports to the Chief Officers of State Library Agencies (COSLA), Anne J. Mathews, Director of the Office of Library Programs, heard similar requests from the nation's state librarians.

In response, Library Programs undertook a project that would determine the current issues and problems in the evaluation of federally funded programs and suggest directions for their improvement. The action agenda included two conferences, a training workshop, a manual on evaluation produced specifically for the state libraries, and the series of White Papers which comprise this publication.

A Midwinter Meeting kicked off efforts in early January 1988. Co-sponsored by the Office of Library Programs and the LSCA Coordinators Discussion Group from the Association of Specialized and Cooperative Library Agencies (ASCLA), a division of the American Library Association (ALA), the intent of the conference was to gather input on the national status of evaluation from state librarians, LSCA coordinators, heads of public library development, and other staff members in state library agencies on several questions:

> What methods and measures are currently used in the assessment of federally funded public library programs?
>
> What are the current issues and problems in evaluation that the state library agencies are facing?
>
> What might the Office of Library Programs do to help improve the process?

These White Papers were commissioned to provide the opportunity for experts in library evaluation to respond to the same questions. A nominal process, conducted with LP's Program Officers, supplied the names of the experts. In the final balloting, those who received the highest tallies were asked to prepare papers. During an organizational meeting possible contributions were discussed and the division of labor set.

Each of the White Papers was peer reviewed by two librarians who were asked to make a judgment about whether the Papers made a contribution to library evaluation by offering new insights and/or by expanding on old ones. Reviewers scored the papers, supplied a critique of the manuscripts, and commented on their suitability for publication. While the authors were asked to cover distinct ground, specifics were left open in the hope that ideas causing ferment in their thinking would make the greatest contribution to the profession and provide the most enlivening reading.

The Papers open with the contribution of Nancy Van House, a general overview of the evaluative process and how program assessment fits within it. After discussing the key elements: How to define effectiveness, develop criteria and indicators, collect data that serve as evidence of effectiveness, and compare current performance with what is desired, she relates evaluative research to experimental design, emphasizing the use of quantitative measurement to ensure validity. Since any evaluation process can be subjective, Van House promotes the use of output measures to make more ends-oriented the evaluation of library program performance.

Douglas L. Zweizig follows by summarizing work to date on adapting Output Measurement to specific library situations. From the concepts and methods used in developing 12 *Output Measures for Public Libraries*, popularized by the 1982 manual and its 1987 revision, he explains and demonstrates how new measures, responsive to local conditions, might be created.

George D'Elia offers reflections on the limitations of output measures. Calling on his past research, he provides us with an ordered analysis of concerns and criticisms, pointing out that program evaluation is driven by a model of the library as a document retrieval system where measurement focuses on outputs. He concludes that it is imperative to expand the scope of currently popularized public library program evaluation by including other functions and by extending the process beyond output measures to include the analysis of outcomes.

Moving from the general perspective, Charles R. McClure turns our attention to specific problems in evaluation based upon the interrelatedness of federal and state library agencies. He compiles and discusses key issues that affect the state library agency's process for evaluating federally funded library programs. Following a review of critical factors for improving the quality of assessment, he details flaws in evaluation at the local, state, and national levels, noting some areas amenable to a relatively quick fix.

Mary K. Chelton also gets to the specifics. After developing the impacts youth services librarians hope to achieve with pre-school, elementary, and secondary school age clients, she explores youth services programs, primarily in public libraries, in terms of their evaluability and offers suggestions for federal funders. The concept of evaluability provides us with a set of procedures for planning evaluations so that stakeholders' interests are taken into account to maximize the utility of the evaluation.

Betty J. Turock asserts that in over 15 years not much new has happened in the evaluation of public library programs, including those that are federally funded. Cautioning that we are wedded to output measurement as we were previously to input measurement, she identifies eight quantitative and qualitative models for assessment that could move us forward in conducting program evaluations.

Leigh Estabrook applies one of these eight models--the old concept of accreditation--to a new institution, the public library. She postulates that accreditation can assist the federal evaluation process by providing information about how a program will benefit from being carried out in the library requesting funds and about how the use of federal funds for a specific program will contribute to the improvement of that library.

Looking ahead, Mary Jo Lynch describes the potential the Federal-State Cooperative System for Public Library Data (FSCS) holds for the evaluation of federally funded public library programs. The emerging database was developed to coordinate the annual Cooperative System for Public Library Data (FSCS) in the evaluation of federally funded public library programs. The emerging database was developed to coordinate the annual collection of public library statistics by state library agencies with the periodic reporting of national public library statistics. In the future it can incorporate information of importance in program evaluation at the national as well as the state level.

Also supplying some new directions, Ellen Altman and Philip M. Clark propose that libraries follow the National Diffusion Network's model for identifying exemplary programs, which in turn could encourage new enthusiasm for improving the process of evaluation.

Taken as a whole, the papers point up the strengths of present day program evaluation and at the same time point up some recurring problems which have, in the main, gone unaddressed--problems that have impact on the assessment of federally funded programs. While crediting the process and products that have brought us this far, they encourage us to return evaluation to an earlier, more inventive momentum, so that we can continue to make progress toward valid, reliable measurement.

These papers continue the work of Ernest R. DeProspo whose 1973 *Performance Measures for Public Libraries*, developed through research sponsored by the U.S. Office of Education, Bureau of Libraries and Learning Resources, served as a landmark which continues to inspire research, debate, and professional growth over 15 years later.

# OUTPUT MEASURES AND THE EVALUATION PROCESS

Nancy A. Van House

## Abstract

Measurement is not an end in itself. It must be understood in the larger context of planning and evaluation. This paper outlines the evaluation process and discusses the role of output measures in the assessment of public library programs. Several key decisions that must be made in evaluation are explored.

Evaluation is an exercise of judgment, an appraisal of value. Public library program evaluation consists of comparing what is with what should be. The evaluation process in its broadest form, therefore, consists of a definition of effectiveness, the identification of criteria and the assessment of the program on those criteria. Output measures are a way of making criteria explicit and measurable. The choice of criteria, however, is complex. Since measurement is not an end in itself, but a means toward more effective public library programs, it must be understood in the larger context of planning and evaluation.

## Effectiveness Defined

Evaluation criteria are derived from the ideal to be achieved, that is, the definition of an effective library program. Most evaluation has defined program effectiveness as goal achievement, explicitly or implicitly (1,2,3,4,5); it is important to realize, however, that this is only one possible definition.

At least four general definitions of organizational effectiveness have been proposed which can be applied to library program evaluation.

The goal or rational system model would define an effective program as one that meets its goals and objectives. This requires that programs have a single set of goals on which participants agree and among which priorities can be set.

The natural systems model would add program health and internal processes to goals. The program must maintain itself as a social system. This model is concerned not just with the ends achieved but with the functioning of the program itself.

The open systems or system resource model would define an effective program as one that acquires from its environment the resources needed to survive. Even goal achievement may become a means toward the end of acquiring more resources. The emphasis in this approach is on the program's relationship with those in its environment who control the resources.

The multiple constituencies or participant satisfaction model is concerned with the extent to which the program meets the diverse, sometimes conflicting, demands of its strategic constituencies. Unlike the goal approach, which assumes that the program has a single set of priorities, this model recognizes that different groups have different priorities. Managers engage in a careful balancing act that requires trade-offs among the preferences of various groups. (6,7,8)

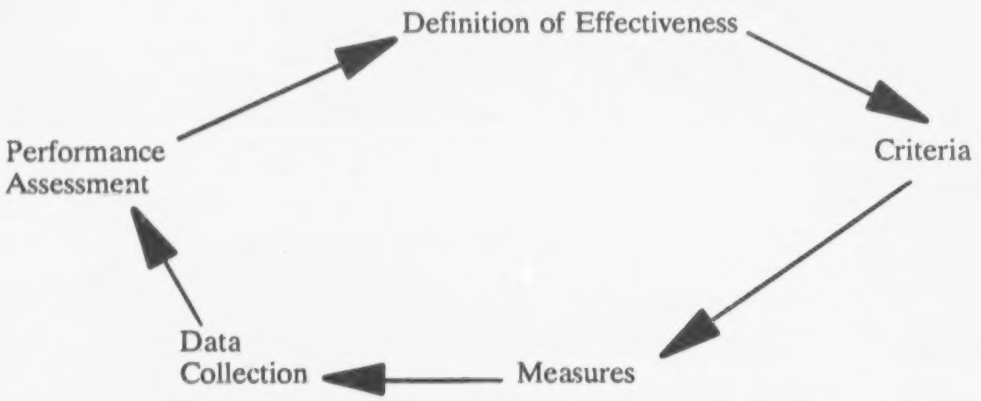
These multiple models of effectiveness are important in that they generate differing criteria for program evaluation. For public libraries, in particular, which function in a political arena, the goal model is useful but limited. Different parts of the community make different demands; budgets are not necessarily tied to services provided; and the people who control the budgets may be responding to their own sets of priorities and political constituencies. While the goal model remains the most useful for program evaluation and the most practical, it is important to realize its limitations.

## A General Model of Library Program Evaluation

The process of library program evaluation begins with the evaluators' definition of program effectiveness and their values and preferences for the library for which criteria for evaluation are developed. Different evaluators may begin with different preferences, which create differences at every succeeding step in the process. If these initial assumptions are not made explicit, however, the source of later disagreements will be unclear at best. These preferences are ultimately subjective, based upon interpretations of what a library program is supposed to do. The goal-setting process is one of reaching consensus on the criteria by which a library program will be evaluated.

Figure 1 presents a general version of the evaluation process adapted from Suchman (9).

FIGURE 1. THE EVALUATION PROCESS



Criteria are abstract. They are made concrete, where possible, by operationalizing them into measures. Measures make explicit and objective the information about program performance used to compare current with desired performance. A criterion for a library program, for example, may be that it serves a target group; measures may include the number of individuals who used the program in the last year and the percentage of users who were members of the target group.

Once the criteria have been developed and measures chosen, the next step is to collect data. The data are compared to expectations in order to judge the effectiveness of the program. This comparison step is crucial: Effectiveness is not determined by the data, but by the judgment process. The same program and the same measurement results may be judged effective by one evaluator, with one set of expectations, and ineffective by another.

The final evaluation step is to close the circle by reflecting on whether values and preferences--and ultimately the definition of effectiveness--should be modified in the light of the outcomes of the evaluation process. Through continual feedback, the criteria, measures, and expectations are cross-checked.

What this model suggests for the evaluation of public library programs is that:

1. Different definitions of program effectiveness are possible, so the choice of a definition of effectiveness and criteria should be made explicit in the evaluation process, generally through the articulation of goals;
2. Measures are needed to make the evaluation criteria concrete. The choice of measures depends on the criteria used;
3. The assessment of program performance depends on the referent against which the measures are compared;
4. The evaluation process is dynamic and cyclical.

## Program Evaluation, Evaluative Research, and Experimental Design

Another way to understand the evaluation process is to relate it to evaluative research (10), which has been important to publicly-funded organizations since the late 1960s. The proliferation of social programs, program-based funding, federal funding, and program planning and budgeting systems (PPBS) spurred the development of reliable methods of program evaluation. Evaluative research uses social science research methods to ask: Did the program cause the desired outcome? This question has two parts: Did the desired outcome occur? If so, can it be attributed to the program or activity being evaluated?

The ideal for evaluative research, as for other kinds of research, is the experiment. Although rarely possible in evaluating library program outcomes, the experimental approach is useful because it bases judgments on evidence and seeks to eliminate alternative explanations for the observed effects. It is basically a skeptical approach, looking not only for results, but for documentation that the program evaluated is responsible for those results.

An experiment is designed to test the impact on subjects of a particular program activity, controlling for other possible causal factors. In its simplest form, it consists of a pretest to assess the initial state of the subjects; the administration of a treatment; and a posttest, to measure the change in the subjects. To reduce the influence of other factors, the experimental group is often matched to a similar control group which receives no treatment, as Figure 2 depicts.

FIGURE 2. THE EXPERIMENTAL APPROACH

| | | | |
|---|---|---|---|
| EXPERIMENTAL GROUP: | Pretest - | treatment - | posttest |
| CONTROL GROUP: | Pretest - | | - posttest |

In evaluative research, the treatment is the program being evaluated. The subjects are the people for whom the program is designed. The tests are the measures of program effectiveness, or impact. In program evaluation, this means that data are needed to assess program impact. Data should be collected before and after the program's implementation, to assess change. If at all possible there should be a control group--a similar program to which the program being evaluated can be compared.

The key concept that carries over from evaluative research to other kinds of evaluation is that of using measurement to determine objectively whether and to what extent the program has had the desired effects. This requires the identification of desired effects, the operationalization of these effects into measures, and the collection of data on the measures. Output measures, especially those found in *Output Measures for Public Libraries*, are frequently used to make key evaluative decisions about public library programs.

**Characterizing Program Evaluation Process**

Several important characteristics of evaluation arose from this discussion:

1. Program evaluation is ultimately subjective. Since each step requires the exercise of judgment, evaluation outcomes depend on whose choices and judgments are applied.

2. The criteria used in evaluation are a function of the program evaluated and the individuals doing the evaluation.

3. Each step in program evaluation is dependent on the steps before it. A lack of agreement among evaluators at one step may result in substantial and increasing divergence at later steps.

4. The more explicit the decisions at each stage the less the likelihood of divergence among evaluators at later stages; or at least the greater the likelihood that the basis for the divergence can be identified.

5. The program evaluation process is made more objective by creating explicit criteria and by using objective data to assess performance.

6. Measurement is an integral part of program evaluation, but in itself is not evaluation. Measurement data divorced from criteria are uninterpretable. In themselves, data provide no information on whether performance is good or bad, only on what it is.

So we can conclude that the program evaluation process consists of the definition of effectiveness; the development of criteria and indicators; the collection of data that serve as evidence of effectiveness; and the comparison of current performance with that which is desired. Ultimately subjective, since it requires the application of values and the exercise of judgment, program evaluation can be moved closer to objectivity by making explicit the decisions that drive it through each step. Measurement data are information that feed the evaluation process, but in themselves they are only raw materials for evaluation decisions.

## Questions To Answer In Assessing Library Program Effectiveness

Cameron and Whetten have presented three major questions that are applicable to libraries when assessing program effectiveness. (11)

### 1. From Whose Perspective Is Program Effectiveness Being Judged?

As noted, the evaluation process is ultimately subjective. Different participants will likely adopt different definitions of effectiveness, use different criteria, and have different expectations for performance on those criteria. Although managers are the most frequent evaluators, theirs is not the only important perspective. Other possibilities for public libraries include users, local government officials, community leaders, library friends groups and trustees, nonusers, library managers, and service providers.

### 2. At What Level of Analysis Is the Evaluation Being Made?

Most public library evaluation takes place at the level of the entire library, the subunit (branch or department), or program (such as Library Services and Construction (LSCA) funded projects). Output measures can be used at any of these levels. The choice of level of analysis drives the data collection process; for example, for some measures the 1987 *Output Measures for Public Libraries* (OMPL) gives instructions for combining branch-level data appropriately in order to aggregate it to library-level data. Simply adding or averaging results across branches or programs that represent different proportions of total library activity can lead to faulty results. Care must be taken to collect and analyze data in a manner that is appropriate to the level of analysis needed.

### 3. What Is the Purpose of Judging Effectiveness?

Some of the purposes of evaluation identified by Weiss include:

To improve program practice and procedures;
To add/drop specific program strategies and techniques;
To institute similar programs elsewhere;
To allocate resources among competing programs;
To continue/discontinue a program; and
To accept or reject a program approach or theory.

Formative and summative assessments serve other purposes. Formative evaluation takes place during the life of a program to gather information for fine-tuning and midstream corrections. Summative evaluation takes place at the end of the program to develop conclusions that may be used for other applications.

### What Is the Time Frame?

Many library program impacts are long term and many are the result of a complex network of interrelated factors, only some of which are under the control of the library and/or attributable to the program being evaluated. Managers often need immediate information to guide their decision-making. The solution may be using more proximate, short-term, means-oriented measures as proxies for long-term, ends-oriented results. The underlying assumption is that there is a causal link between short-term means and long-term ends.

Output measures generally present a snapshot of current activity. They are most useful, however, when used repeatedly, since repeated measurements show changes and trends over time. Long term data are even more helpful. For example, a program to increase library use among a target group may be judged successful if before-and-after data showed an increase. However, even earlier data may show that there was an increasing trend in use among the target group before the program was implemented; in that case, to be judged successful the program should have resulted in an increase greater than would have been expected from the trend observed. On the other hand, later data, collected well after the end of the program, may show that the increase resulting from the program was only temporary.

### What Is the Referent Against Which Effectiveness Is Judged?

In comparing what is with what should be, decisions have to be made about the program's expected performance. Jane Robbins and Douglas Zweizig have provided us with an excellent variety of referents in their continuing education series on evaluation, published by *American Libraries* in 1985. (12)

### What Types of Measures Are Used for Judgments of Effectiveness?

Five categories in addition to output measures are commonly employed:

Effort: Program inputs, including the quantity and quality of activity.

Output or performance: The results of effort, including the quantity and quality of services or programs, and the number of people served.

Adequacy of performance: The degree to which performance is adequate to the total amount of need.

Efficiency: The ratio of effort to performance, inputs to outputs.

Process: Internal operations, including numbers of activities carried out by the program that are means rather than ends.

Outcomes: Effects on persons served; effects on the organization; effects on larger systems, including networks of agencies and classes of organizations; effects on the public, including changes in public values or attitudes.

All of these may be used in evaluation. Libraries, although turning more and more to output measures, traditionally collect data on inputs, but input data reveals nothing about how appropriate the resources were or how well they were used. Adequate performance is difficult to measure. Efficiency and process measures are useful primarily to diagnose internal organizational performance. Measures of outcomes or effects, while highly desirable, are difficult, if not impossible. For example, we may know that many children use the library, but we cannot measure its impact on their lives.

Measures may be objective or subjective. Objective measures consist of data that directly reflect performance, for example, circulation or number of reference transactions. Subjective measures represent someone's assessment of performance, such as a librarian's rating of the adequacy of program performance on circulation of materials or reference service.

Output measures, now widely promulgated for library evaluation through the first and second editions of *Output Measures for Public Libraries* (OMPL), are primarily objective, but both subjective and objective measures have a role in evaluation. Some aspects of performance are best measured subjectively, for example, user perception of librarians' helpfulness. In some cases, objective measurement may be possible but difficult and subjective measures may be a satisfactory approximation, e.g., librarian assessment of reference completion rate, as opposed to an objective test of whether each question was answered.

To be useful for evaluation, measures must be pertinent, reliable, valid, sensitive, and feasible.

Pertinent measures are relevant to the criteria being used for evaluation.

Reliable measures give the same results with repeated application, as long as the program characteristics or behavior being measured have not changed. Changes in the measurement results are due to changes in the program, not to variations in measurement. Comparisons can be made across programs or over time only if the data have been collected in the same way each time.

Valid measures accurately represent the criteria to be evaluated. Invalid measures often reflect a closely-related concept, but not the actual criterion being applied.

Sensitive measures respond to changes in program characteristics or behavior. Insensitive measures require large changes in performance before differences show up in the data.

Feasible measures are within the capability of the library. They do not require extraordinary resources nor are they excessively limited in the circumstances under which they can be used.

## Using Output Measures To Evaluate Public Library Programs

Key sources of library output measures have developed over the last 15 years through the work of DeProspo and others (13), Kantor (14), Lancaster (15,16), Van House and others (4, 17), and Zweizig and Rodger (18). The use of output measures has been an attempt to make more objective and ends-oriented the evaluation of library performance. Stimulated in part by the scarcity of library resources and the need to maximize the benefits of the resources available, it has been accompanied by a goal-oriented approach to planning and evaluation. Output measures have also sought to take the users' point of view.

Used for internal management decision making and for external justification of the library's need for resources, output measures reflect the extent and effectiveness of library programs by focusing on the results of program activity as well as the quantity and quality of services delivered.

Output measurement for evaluation has encouraged libraries to make objective the process by which they evaluate programs and make decisions; to demonstrate the effectiveness of programs internally and externally; to communicate their service orientation to external constituents; and to base resource allocations on objective data. But these measures must be interpreted in the light of their role in the evaluation process. Whether the results of a library program's output measures are acceptable depends on a number of factors, including the purposes for which the evaluation is being done, the goals of the program, the needs and preferences of the community being served, and the resources available. Another major factor is who is doing the evaluation and whose judgment prevails. This is ultimately a political decision from which no measurement methods or results will insulate library program managers.

Information is still lacking about the determinants of output measures, that is, the relationship between library resources and activities and community characteristics, on the one hand, and output measurement results. Library managers continually make decisions aimed at improving their library's performance, but at present little empirical data are available to guide those decisions. This is an important point, because until we better understand the relationship between causes of output measures and their results, we must be careful not to over-interpret their meaning.

**Notes and References**

1. Orr, R.H. "Measuring the Goodness of Library Services: A General Framework for Considering Quantitative Measures," *Journal of Documentation* 29 (September 1973): 315-332.

2. Du Mont, Rosemary Ruhig, and Paul F. Du Mont. "Assessing the Effectiveness of Library Service," *Occasional Papers, 152* (Champaign: University of Illinois, Graduate School of Library and Information Science, 1981).

3. McClure, Charles R.; Amy Owen; Douglas L. Zweizig; Mary Jo Lynch; and Nancy A. Van House. Planning and Role Setting for Public Libraries: A Manual of Options and Procedures (Chicago: American Library Association, 1987).

4. Van House, Nancy A.; Beth Weil; and Charles R. McClure. *Output Measures for Academic and Research Libraries* (Chicago: American Library Association, 1989) in preparation.

5. Childers, Thomas. "Evaluative Research in the Library and Information Field," *Library Trends* 38 (Fall 1990), in preparation.

6. Cameron, Kim. "Domains of Program Effectiveness in Colleges and Universities," *Academy of Management Journal* 24 (March 1981): 25-47.

7. Scott, W. Richard. *Organizations: Rational, Natural, and Open Systems* (Englewood Cliffs, NJ: Prentice-Hall, 1987).

8. Childers, Thomas, and Nancy Van House. "Dimensions of Public Library Effectiveness," *Library and Information Science Research* 11 (October-December 1989) in preparation.

9. Suchman, Edward A. *Evaluative Research: Principles and Practices in Public Service and Social Action Programs* (New York: Russell Sage Foundation, 1967).

10. Weiss, Carol H. *Evaluative Research* (Englewood Cliffs, NJ: Prentice-Hall, 1972).

11. Cameron, Kim S. and David A. Whetten. "Some Conclusions about Organizational Effectiveness." In: Cameron, Kim S. and Davis A. Whetten, Editors. *Organizational Effectiveness: A Comparison of Multiple Models* (New York: Academic Press, 1983): 261-275.

12. Robbins-Carter, Jane and Douglas L. Zweizig. "Are We There Yet?" *American Libraries* 16 (October 1985): 624-627; 16 (November 1985): 724-727; 16 (December 1985): 780-784; 17 (January 1986): 32-36; 17 (February 1986): 108-112.

13. DeProspo, Ernest R.; Ellen Altman; and Kenneth E. Beasley. *Performance Measures for Public Libraries* (Chicago: American Library Association, 1973).

14. Kantor, Paul B. *Objective Performance Measures for Academic and Research Libraries* (Washington, D.C.: Association of Research Libraries, 1984).

15. Lancaster, F.W. *If You Want to Evaluate Your Library...* (Champaign: University of Illinois, Graduate School of Library and Information Science, 1988).

16. Lancaster, F.W. *The Measurement and Evaluation of Library Services* (Washington, D.C.: Information Resources Press, 1977).

17. Van House, Nancy A.; Mary Jo Lynch; Charles R. McClure; Douglas L. Zweizig; and Eleanor Jo Rodger. *Output Measures for Public Libraries: a Manual of Standardized Procedures*, Second edition (Chicago: American Library Association, 1987).

18. Zweizig, Douglas and Eleanor Jo Rodger. *Output Measures for Public Libraries* (Chicago: American Library Association, 1982).

# ADAPTING OUTPUT MEASURES TO PROGRAM EVALUATION

**Douglas L. Zweizig**

## Abstract

*Output Measures for Public Libraries* presents 12 measures that relate to commonly occurring public library objectives. The concepts and methods used to develop these 12 can be applied to the design of additional measures specific to library program evaluation. Set within a goal orientation, candidate measures are developed and explained.

Since the landmark publication of *Performance Measures for Public Libraries* in 1973 (1) the use of output measures as a means of evaluation has captured professional attention. With the release of *Output Measures for Public Libraries*, first and second editions, by the Public Library Association (PLA) in the 1980s (2, 3), interest in the use of output measurement has grown steadily. Numerous state library agencies have incorporated them in regular data collection and in public library standards. Along with familiarity has come the awareness that, while the basic set of measures may have universal applicability for public libraries, it may not apply to particular libraries and particular programs, such as those funded under the Library Services and Construction Act (LSCA). The approach, however, can be used to develop measures that more closely match the needs of particular program situations.

## Output Measures Features

Output measures are intended to be standardized indicators of what a library gives to its community. The latest PLA output manual provides definitions of the elements to be measured and specifies procedures for gathering data. The operational instructions support comparability. Obviously, full comparability is an ideal; individuals bring their own needs and interpretations to the data collected, but without an emphasis on standardized procedures no comparison is possible. By definition, a measure designed for a local situation or particular program is not intended as a generally comparable measure. Still the need for standardization remains, if data are to be compared from one program to another, from one year to another, or from one library locale to another.

Output measures are designed to be affordable, to be capable of use by small as well as large libraries. When original decisions were made about the design of an output measure, the measure with the least effort was chosen if it would provide information that was considered "good enough." Measures with more complexity were discarded. Library programs have limited budgets for evaluation activities. The purpose of designing less expensive measures was to provide methods that were so affordable that they would be used widely and with some frequency. Since the point of output measurement is to provide data that will assist program administrators in their decisions, it is important that current data on performance is obtained easily. More elaborate or more costly measures are less likely to be used frequently, leaving large periods of time for which no useful data are available.

The results of output measures are easily communicated to staff, Boards of Trustees, public funders, and other potentially interested persons. As far as possible, jargon and the use of data tables were avoided in the design of the measures. It is possible to express the results of an output measure in a simple sentence, without even using the name of the measure. For example, the results of the measure for circulation per capita for a program's target population can be described with the statement, "On the average, 11 percent of the elder target audience borrowed 6,000 books from the library last year." Data from output measures can be used in narrative reports and in newspaper stories to express the contribution of the program to the community.

Output measures are ratios, composed of data elements. For example, circulation per capita is a ratio formed by dividing the annual circulation by the population of the legal service area. Many of the measures are expressed in terms of population figures since the size of the population to be served clearly impacts on the amount of service provided. For the measures of materials availability, the number of successes is divided by the number of attempts to obtain a percentage of success.

Output manual instructions for data gathering are for the collection of data elements. The final output measure is produced by dividing one data element into another. So if a new measure is to be a ratio, close attention needs to be given to how the data elements are to be obtained. Further, the same data element can be used in more than one output measure. For example, population of the legal service area is used in 6 of 12 measures. The use of a data element in more than one measure reduces the cost of data collection. For program evaluation, a data element that would have repeated adaptations would be the population of the target audience.

Another of the ways in which reductions in levels of effort have been introduced into measures of effectiveness is in the use of sampling. For many libraries this has been an innovation. In statistics collection, librarians are accustomed to counting everything under the mistaken belief that counting everything produces more accurate data. Some reflection would indicate that there are accuracy problems with trying to count everything. The practice of counting reference statistics is an example. The usual method is to make hash marks on a form to indicate each reference transaction. But program staff are busy; marks are not made after each transaction. Then, when they have a moment free, the staff try to recall how busy they were and put down a group of marks to represent their memory. It would be hard to see these marks as an accurate count of the actual number of program transactions.

It is equally important that what is being counted is clearly defined. Many programs record reference transactions separately from directional transactions. Over the year, staff members' definitions of this distinction begin to waiver and blur. Different staff members end up recording different things as reference transactions and their hash marks are added together, even though the hash marks record differences. When each reference transaction is being counted for the entire year, any inaccuracies in today's data are simply added to the year-to-date figure, a figure already flawed by earlier error. At the end of the year, a number is produced that summarized the hash marks made in 1989. We know that the number is not accurate, but we have no way of knowing how inaccurate it might be.

If a library uses sampling -- if it records its program's reference statistics for only a week or two -- the library staff can give quality attention to the data collection for a limited period of time. In a larger library situation, a staff member can be designated to supervise the data collection, can observe whether staff are indeed counting transactions properly, and can arbitrate decisions about whether a given transaction should be counted as a reference transaction or not. So in a limited period of time, high quality data are collected and the program's annual reference transaction count can be estimated from that sample. Information on how to design samples and on selecting the size of sample needed are given in *Output Measures for Public Libraries*, second edition. (4) The advantages of sampling for program evaluation -- more accurate data, more affordable data collection, and quicker answers to questions -- make total counts seem highly impractical.

Output measures can provide the basis for answering the key evaluation question: "Compared to what?" Evaluation is often treated as if the answer to that question were obvious, but it rarely is. Different schools of thought rely on different bases for comparison, giving us a selection of "compared to what's" that can be used.

We often compare our program performance today with that in the past. "Are we doing better this year than last year? Are we circulating more self-help books, answering more health questions..." So one standard for comparison is what the library was doing in the past, or what the library was doing before the program was inaugurated.

Published library standards are another means for program comparison. When the library association or state library tells librarians that they ought to be providing a certain level of service, they can match program performance against the level specified in the standard.

A third standard for comparison is the performance of other libraries. One way librarians make decisions about their programs is to ask whether they are doing better or worse than the neighboring or comparable libraries.

The performance of a program can be compared with expectations. The key question here is, of course, "Whose expectations?" One group of evaluation proponents, for example, is concerned with stakeholder analysis. They identify the groups interested in the program's evaluation and determine their varying expectations of performance.

Some evaluators believe that a library program should be assessed in terms of survival or growth. If the budget keeps going up, if the program survives, then it must be doing well.

But the standard for comparison that seems most meaningful is, "What was the library program trying to do?" Rather than what state standards say the program should be doing, what the program in a neighboring library is doing, what the program did last year, what the program was trying to do requires integrating evaluation with goal-setting.

Since output measures are designed to provide evidence of achievements toward commonly occurring program goals of public libraries, it is important to understand which output measures relate to these goals and the data elements that make up each of the output measures. Some examples are depicted in Figure 1.

The common goals, shown in the initial line are followed by the measures in the second line and the data elements which comprise the measures in the third. Linking goal areas and measurement is a central point when the standard for evaluation of performance is what was being attempted. If a library program has chosen the goal of maximizing the use of materials for a target population, then it will have some interest in its performance for the target program population on the associated measures of circulation per capita, in-library materials use per capita, and turnover rate.

## FIGURE 1. GOAL AREAS AND RELATED OUTPUT MEASURES

**REACHING MAXIMUM NUMBER OF PEOPLE IN TARGET AUDIENCE** (Library Use Measures)

**LIBRARY VISITS PER CAPITA OF THE TARGET AUDIENCE**

(Annual number of library visits by target audience / Potential target audience from population of legal service area)

**TARGET AUDIENCE REGISTRATION AS A PERCENTAGE OF THE TARGET POPULATION**

(Target audience registration / Potential target audience from population of legal service area)

---

**MAXIMIZING USE OF MATERIALS** (Materials Use Measures)

**CIRCULATION PER CAPITA FOR THE TARGET AUDIENCE**

(Annual target audience circulation / Potential target audience from population of legal service area)

**IN-LIBRARY MATERIALS USE PER CAPITA OF THE TARGET AUDIENCE**

(Annual in-library use of the target population / Potential target audience from population of legal service area)

**TURNOVER RATE**

(Annual circulation of target collection / Total holdings in target collection)

---

**PROVIDING READY ACCESS TO MATERIALS** (Materials Access Measures)

**TITLE FILL RATE FOR TARGET AUDIENCE**

(Number of titles found by target audience / Number of titles sought by target audience)

**SUBJECT AND AUTHOR FILL RATE FOR TARGET AUDIENCE**

(Number of subjects and authors found by target audience / Number of subjects and authors sought by target audience)

### BROWSER FILL RATE FOR TARGET AUDIENCE

(Number of browsers from the target audience finding something / Total number of browsers from the target audience)

### DOCUMENT DELIVERY TO TARGET AUDIENCE

(Number of materials available to the target audience within 7 days / Number of requests from the target audience for materials not immediately available)

(Number of materials available to the target audience within 14 days / Number of requests from the target audience for materials not immediately available)

(Number of materials available to the target audience within 30 days / Number of requests from the target audience for materials not immediately available)

## PROVIDING INFORMATION IN RESPONSE TO QUERIES FROM TARGET AUDIENCE

(Reference Services)

### REFERENCE TRANSACTIONS PER CAPITA FOR THE TARGET AUDIENCE

(Annual number of reference transactions from target audience / Potential target audience from population of legal service area)

### REFERENCE COMPLETION RATE FOR THE TARGET AUDIENCE

(Number of completed reference transactions from target audience / Number of reference transactions from target audience)

## PROVIDING INFORMATION PROGRAMMING FOR THE TARGET AUDIENCE

(Programming)

### PROGRAM ATTENDANCE PER CAPITA FOR THE TARGET AUDIENCE

(Annual program attendance of the target audience / Potential target audience from population of legal service area)

For many public libraries, this set of goals and associated measures will encompass the program's intentions and achievements. To best use output measures for evaluation, the library staff should set objectives for library programs in the goal areas most important for it. For example, "To increase circulation per capita for the target population from (present score) to (desired score) by June 30, 1990." The calculation of the score for circulation per capita on June 30, 1990 will reveal how close program performance came to the desired level of performance.

These characteristics of output measures as presented in *Output Measures for Public Libraries*, second edition -- standardized indicators, affordable, easily communicated, composed of ratios, using sampling, and focused on objectives -- can be designed into new measures that evaluate other aspects of library program performance.

## Fitting Output Measures to an Overall Evaluation Process

The purpose of evaluation is not just to know whether to feel good about some aspect of a library's program. Its purpose is to allow us to make better decisions about the program -- to identify aspects that might be improved. Rather than asking, "How good is it?" the evaluation approach suggested here asks, "Are we there yet?" -- a question that better reflects the uses to be made of the evaluation. (5) This question sees evaluation as a process of checking regularly to determine how much progress has been made towards a stated program goal.

The question, "Are we there yet?" grows out of a working definition of planning, which is a series of successive approximations towards a moving target. That is, planning involves a repeating process through which the library program moves closer to its intended goal, the target. However, while the library is carrying out its plan, the demands being made on the program and the characteristics of its environment change; the target moves even as the library attempts to approach it. Assume, for example, that a library's target for a collection development program, partially funded under Title I of the Library Services and Construction Act (LSCA), is to achieve a title fill rate of 70%. If the library's annual materials budget is cut, the target of 70% may no longer be appropriate and may have to be changed. Planning's role is to select the target and to review whether that target has moved. Evaluation's place in the planning process is to periodically assess how much closer the library program's performance is to the target.

## A Seven Step Process

For the sake of clarity, the evaluation process has been separated into seven steps, but it is not an elaborate process. The whole record could be transcribed on a half sheet of paper, as we see in Figure 2.

### 1. Determine the target area.

The target area can refer to what you want the program to accomplish (effectiveness) or to how well you want the program to do it (efficiency). The process is very similar to determining a goal in planning. For example, a target area for a library program may be how much the materials are being used by a specific audience or how many of that audience are registered at the library. A helpful question for determining the target area is, "What specifically do you want to know about it?" The answer may be that you want to know about several things: Percentage of increase in materials used by the target audience, percentage of increase in the target audience registered, and the ratio of the target audience registered to the potential target audience. In such a case, each of these is a separate target area. A description of each target area should be listed on a separate Evaluation Summary Sheet.

FIGURE 2. THE SEVEN STEP EVALUATION PROCESS

**EVALUATION SUMMARY SHEET**

o DETERMINE TARGET AREA:

| Target | Actual | Difference |
|---|---|---|
| | | |

o HOW WILL YOU KNOW? (procedures for collecting data):

o SO WHAT?:

o RETHINKING DECISIONS:

2. Set the target.

For each target area, a specific target needs to be set. That is, each target area identifies the aspect of the program you want to evaluate, but the target itself is the specific standard you will compare your results against. The target should be measurable (i.e. expressed in a number), or at least observable (i.e., you should be able to tell unambiguously whether you've done it or not). Examples of measurable program targets are: To increase registered older adults 30 percent by June 30, 1990; to increase the annual number of in-service/continuing education hours per staff in providing services for older adults from 2 to 3 by December 31, 1990. For each target area, a number of possible targets should be considered and a selection made from the set. The process of considering alternative measures will clarify many measurement issues.

The target should be entered on the Evaluation Summary. These measurable targets are the same as the objectives described in *Planning and Role Setting for Public Libraries.* (6)

## 3. How will you know?

*Output Measures for Public Libraries,* second edition, suggests specific target areas appropriate for public libraries. If these target areas are applicable to a specific library program, then the procedures in the manual can be used for collection of the data that will tell whether the program has met the target.

If the library selects target areas other than those provided in the manual, then procedures for obtaining needed data will have to be developed. Sometimes the information that tells whether the target has been met is obvious. At other times, decisions must be made about the data required to determine how close performance has come to the target. For example, how will a library know whether its efforts to recruit more volunteers for services to the handicapped and shut-in have been successful?

If procedures are needed for the collection of relevant data, this is the place to spell them out. For example, at the end of December 1989, the number of volunteers recruited in the year will be recorded and will be divided by the number recruited in 1988; this number will be multiplied by 100 and then 100 will be subtracted from it to produce the percentage of increase or decrease in volunteers recruited.

So, if you recruited 20 volunteers in 1989
and had recruited 15 in 1988:

20 divided by 15 = 1.33,
times 100 = 133,
subtracting 100 = +33%

If you recruited 15 volunteers in 1989
and had recruited 20 in 1988:

15 divided by 20 = .75,
times 100 = 75,
subtracting 100 = -25%

As the evaluation is designed, the procedures for gathering the needed data should be entered on the Evaluation Summary.

## 4. Take a look.

During this step, the evidence needed for the evaluation is gathered to produce the figure that corresponds to the target: The percentage increase in older adults registered; the percentage increase in volunteers recruited. The actual achievement should be entered on the Evaluation Summary.

5. How close are you?

At this point, the comparison between the target and the actual figure is made. Judgment should be suspended and only the facts recorded. The difference between the target and actual performance is usually calculated by subtracting the target from the actual figure, so that if the actual is greater than the target, the difference will be positive, and if the actual is less than the target, the difference will be negative.

6. So what?

When the data on actual performance and the target have been recorded together, someone needs to make a decision about whether to act on any difference between them and what to do. If the actual growth in volunteers is 10% and the target was a 20% increase, several kinds of decisions are possible: Increasing recruiting efforts; enlisting the assistance of the existing volunteer organization; or setting the target at a different level, if the original level did not anticipate a change in the availability of potential volunteers. The decisions should be recorded on the Evaluation Summary Sheet.

7. Rethink.

Each step in the evaluation process may result in learning more about the aspect of the library program being evaluated. The program may have a target area that is too difficult or too expensive to measure and it may need refinement. If the library does not have enough information to set a target, it may need to collect some data before setting one. Changes that occurred after the target was set may require rethinking the target.

At any stage of the evaluation, a library program manager may decide to go back and fix up some earlier step. This procedure is not only appropriate, it is necessary if the evaluation is to be useful. In recording the process and results of the evaluation, however, changes that resulted from rethinking should also be recorded so that anyone interested in the evaluation can determine what you've done.

Finally, although such a step is beyond the evaluation process itself, it is important to communicate results. Staff need to know the library program targets and how the library is doing in reaching them. Evaluation results can highlight in specific figures for the tight-fisted funder the demands made on the library and the success programs have in meeting such demands. The results can point out program progress in increasing use, areas where staff are overextended, or areas where the equipment or materials are insufficient to respond to demand. In short, the purpose of evaluation is to enable the program to operate better in the future by identifying areas needing improved performance or increased resources.

Developing specific program measures fits into an overall goal-setting and evaluation process. Going through this process will focus attention on the purpose of the measures and the decisions that will be based on its results.

## Criteria for Selecting a Measure

Earlier it was recommended that a number of alternative measures should be considered for each target. A series of criteria can be applied to evaluate those alternative measures.

EASE: How easy would it be to obtain this measurement? Some data are already available or are easy to collect; others require special studies. Where choice is possible, the easier measure is preferred, not only because it is less costly, but also because it is more likely to be collected accurately and as often as needed.

INTRUSIVENESS: How intrusive would data collection be? Some data can be collected without the users' or clients' awareness; others can be collected only with the cooperation of the user or they will cause a disruption in the program.

MEANINGFULNESS: How meaningful would the results be to the public officials who funded the program? Some data may be of interest to the staff, but may not speak to the interests of those who provide the funds for the program. This might be called the recognition factor; in research parlance, this is known as face validity. That is, does the measure look like the right measure for what you're interested in, will people accept it as a measure?

COMPARABILITY: How comparable would the results be with those from other similar programs in similar libraries? If different libraries use different definitions of terms, the results of their data collection may not be comparable. For some measures, comparability with different libraries may not be desired, but the issue of comparability within the library and across time will still need to be addressed.

VALIDITY: How close is the measure to assessing what the library really wants to know about the program? It's possible to be measuring something other than what was intended. For example, early studies of voter behavior would conduct door-to-door surveys and ask, "For whom did you vote in the last election?" When the researchers tallied the results, their data showed that many more people said they had voted than had actually been recorded at the polls. The problem was that people were reluctant to admit to the interviewer that they had not voted. So the question was not measuring voter behavior; it was measuring people's desire to be seen as good citizens. When the surveyors used a screening question, "A lot of people weren't able to get to the polls for the last election. Were you able to get to the polls?" Those who answered, "Yes," could be asked about their choices in the voting booth, and the results more closely reflect actual voting behavior.

When a waiter comes by your table and asks, "Is everything satisfactory with your meal?" your response is probably not a measure of the quality of the meal. Your quick, kind words are more likely a measure of your reluctance to interrupt your meal at that moment. Similarly, our measures of user satisfaction can fail to measure the quality of our service and tap instead the politeness of our target users. A more valid measure would ask, "In what way did the program help you?" or would focus more on behavior than on attitude, "How were the results of our program used?"

RELIABILITY: To what degree are different program staff likely to collect data in the same way? Some measures will be easier for staff to record accurately; the understanding of other measures will differ from one staff member to another.

One of the things that helps a sailor navigate is a buoy. We anchor an object that floats and we steer by it; it's a reference point. A valid buoy is one that is anchored directly over the desired mark or over the hazard. A buoy will lack validity to the degree that it is not indicating the underwater condition it is supposed to mark. Reliability is looking at the length or elasticity of the line connecting the buoy to the anchor. If the line is too long or too springy, the buoy is able to move around on the surface and will be an unreliable indicator. With a shorter line, the location of the buoy will be fairly stable and will be a more reliable indicator.

A point to be made here is that validity or reliability of a given measure is not absolute; it is variable. It is not meaningful to say a measure is not valid or not reliable. The question is, "How valid is it? How reliable is it?" And these questions are relevant to the library program manager because the validity and reliability of a measure are directly related to cost. The evaluator can determine how much reliability or validity a measure has, but it is the program manager who needs to decide how much of each to buy, how much is necessary for the decisions the manager needs to make. For example, unobtrusive measurement of reference services is believed to be a more valid measure of quality of service than asking the reference staff to indicate which questions are answered satisfactorily, but testing is more complex and costly. Personal interviews cost at least ten times more than mail surveys; are they worth it? Our estimate from a sample can be twice as accurate if we quadruple the sample size, how good an estimate do we need? These are all questions that the program manager needs to address when making decisions about measuring program effectiveness.

The reliability of data collected from a sample is directly related to the size of the sample. The exception to this is when the population being sampled is small -- below 2,000; almost all library program data are from larger populations. When national polls are conducted, generally samples of about 1,500 people are used to estimate national characteristics. These estimates are found to be within 2% of the national figure (95% of the time). *Output Measures for Public Libraries* (1982) recommended small sample sizes -- intentionally reduced reliability -- in order to keep the costs of measurement low. The sample sizes recommended provide estimates within 10% of what would be found if everything (every reference question, every library visit) were counted with the same accuracy. We made this decision because we felt that library program managers were not interested in small differences and that closer estimates would not be cost effective. The sample size needed to produce results within 2% is about 16 times the sample size needed to produce results within 10%, so the program manager needs to weigh the costs of reliability.

CONTROLLABILITY: To what degree can the program control the outcome for the measure? Some of the program's impact can be controlled to some degree by the library; some of the impact is primarily in the user's control. There seems to be an inevitable tension between the aspects of service of most interest and the aspects over which the library has control. The library program intends that its materials will help users better understand themselves and the world around them, but the program staff have virtually no control over the uses made of the materials. On the other hand, the program has a large amount of control over which materials are owned. In between these extremes, the program shares control with the user: The library program can control to a greater extent which materials circulate, to lesser extent whether material that is circulated is read. The most satisfactory output measures are those that not only point toward desired impacts of service but that also can be affected by management decision making -- measures in the middle range, such as circulation or demand for service.

GOAL-RELATEDNESS: To what extent does this measure relate to important program goals? Some measures will be directly related to key goals, or target areas, of the program of service; others will be related to peripheral goals. Being sure that the measures used relate to an important program goal will help ensure that the measurement data can and will be used for planning and evaluation.

## Adapting Outputs to Measure Programs

The concepts and methods of output measures can be adapted to aspects of library programs not addressed in the original 12 measures. In selecting new aspects for measurement, a goal-oriented approach is strongly recommended. Criteria such as ease, intrusiveness, meaningfulness, comparability, validity, reliability, controllability, and goal-relatedness can help to assess the degree to which the adapted measures can provide cost-effective and useful data.

## Notes and References

1. DeProspo, Ernest R.; Ellen Altman; and Kenneth E. Beasley. *Performance Measures for Public Libraries* (Chicago: American Library Association, 1973).

2. Zweizig, Douglas L. and Eleanor Jo Rodger. *Output Measures for Public Libraries: A Manual of Standardized Procedures* (Chicago: American Library Association, 1982).

3. Van House, Nancy; Mary Jo Lynch; Charles R. McClure; Douglas L. Zweizig; and Eleanor Jo Rodger. *Output Measures for Public Libraries: A Manual of Standardized Procedures,* second edition (Chicago: American Library Association, 1987).

4. Van House: 18-22.

5. Robbins, Jane and Douglas L. Zweizig. *Are We There Yet? Evaluating Library Collections, Reference Services, Programs, and Personnel* (Madison: School of Library and Information Studies, University of Wisconsin-Madison, 1988).

6. McClure, Charles R.; Amy Owen; Douglas L. Zweizig; Mary Jo Lynch; and Nancy Van House. *Planning and Role Setting for Public Libraries: A Manual of Options and Procedures* (Chicago: American Library Association, 1987): 46-49.

# BEYOND OUTPUT MEASURES

**George D'Elia**

## Abstract

The evaluation of library services is currently driven by a model of the library as a document retrieval system where measurement focuses on outputs. While the preeminence of the document retrieval function is recognized and while the measurement and evaluation of outputs are clearly necessary, it is imperative to expand the scope of the evaluation of public library services both by including other functions and by extending the process beyond output measures to include analysis of outcomes. A case study is presented in which evaluation is expanded by assessing a library in terms of the roles it plays in the community.

The framework for evaluating human service programs generally consists of three aspects: 1) The community problem which the program addresses; 2) The set of activities which are performed by the program; and 3) The measures for assessing the degree to which these activities have achieved their intended outcomes. (1) Implicit in this framework is the expectation that the cumulative effects of the programmatic activities will have the desired impacts on the community problem of interest.

## A Conceptual Framework for Evaluating Human Service Programs

Within this three-pronged conceptual frame a problem is defined as an unsatisfactory condition prevailing in the community served by the agency. It is assumed that this condition will remain unsatisfactory without the intervention of the program. The determination of an unsatisfactory condition is a subjective process that is dependent on the community's social values, the clarity of the mission of the agency, and the compact between the agency and the community which it serves. An unsatisfactory condition in the community is not considered a problem in relation to the agency unless an anticipated program or service it can offer will affect the condition.

The purpose of the program is the amelioration of the unsatisfactory condition, which is designated the inherent problem; the desired outcome of the program is designated the inherently valued outcome. Ideally, both are amenable to objective measurement; the difference between the two measures represents the impact of the program.

In order to achieve the inherently valued outcome, the program identifies a set of intermediate outcomes to be achieved and undertakes a series of activities designed to achieve them. In effect, the program is a presumed causal chain of activities and outcomes. The program articulates what is to be done and why, what is to be accomplished, and how it is to be accomplished.

## Framework for the Evaluation of Public Library Services

Evaluations of public library services tend to be executed at two levels--the general program level and the specific program level. At the general program level, library services are evaluated which have been designed to meet broadly defined community needs. Given that these services, such as reference and circulation, tend to be common to all libraries, the procedures for executing them can be standardized and can yield data that permit comparative assessments among libraries. However, it is important to bridge the gap between general program assessments and those for specific programs, since procedures applicable at one level can be appropriate at the other.

At this time, the procedures for evaluating library services can be characterized as: 1) driven by a model of the library as a document/ information retrieval system; and 2) focused on the outputs of the library's activities.

## The Library as a Document Retrieval System

Proponents of the document retrieval system model maintain that libraries exist to bring documents and users together. (2,3) The library operates as, "an interface between the available information resources and the community of users to be served. Therefore, any evaluation applied to the library should be concerned with determining to what extent it successfully fulfills this interface role". (4) In this model, the outcome of library services is document/information delivery. The programmatic activities antecedent to this outcome include the acquisition, organization, storage, retrieval, and dissemination of documents and information. An extensive body of procedures for evaluating these activities has been developed. These procedures, as reviewed and synthesized by F. Wilfrid Lancaster, represent a formidable professional accomplishment and provide ample approaches for those interested in evaluating any programmatic activities.

While the primacy of the document retrieval function of libraries is evident, it is also evident that public libraries perform other functions in society which need addressing.

## Criteria for Evaluating the Library and Its Activities

Lancaster describes the evaluation of library services in terms of inputs, outputs and outcomes. The inputs are the materials and resources that are employed by the library to perform activities or to provide services. The outputs are the quantifiable indicators of the activities which are performed or the services which are provided. Inputs can be evaluated in terms of their contribution to the achievement of the desired outputs. Outputs, in turn, can be evaluated either in terms of the attainment of some desired level, which is inherently valued, or by comparison with levels achieved by other libraries. Success in achieving the desired level of output for given activities is generally assumed to be a measure of library performance as is a library's standing relative to the outputs of other libraries.

The outcomes of library service represent the desired impacts of that service on the community. They tend to relate to long-term social and behavioral objectives, such as, "improved level of education, better use of leisure time, and a more aware and socially responsible citizenry." (5) Lancaster notes that these outcomes tend to be intangible, not easily converted into measurable evaluation criteria and, "too vague and impractical to be used as criteria by which one can readily evaluate a library or its services." (6) While these outcomes provide the justification for the existence of the library, he notes that it is virtually impossible to measure the degree to which they have been achieved and, even if the measurements were possible, it would be difficult to isolate the contribution of the services to the desired outcomes. For these reasons, Lancaster suggests abandoning the idea of using desired outcomes as direct criteria for the evaluation of libraries and suggests instead focusing on the library's outputs. However, while the identification and measurement of the outcomes of library services are difficult, considerable benefits are accrued by expanding the scope of library evaluations to include them.

## Evaluating Output Measures

Current procedures, developed and promoted by the Public Library Association for evaluating public library services, are based on the measurement and evaluation of outputs as presented in *Output Measures for Public Libraries.* (7) This manual identifies 12 output measures which are purported to be valid, reliable and, because of the standardized procedures prescribed in the manual, comparable across libraries. While these efforts have been successful in focusing the profession's attention on evaluation, the measures should be used cautiously.

Seven of them are designed to measure extent of use. While it is self-evident that the library exists to be used and that use of the library's materials, services and programs is a tangible indicator of community demand, use may be engendered by social processes that are beyond the control of the library more than by the quality of the library's performance.

Five of these output measures purport to assess the performance of the library as a document/information retrieval system which is but one of the missions of the public library in society. In addition, some of these measures are demonstrably invalid or unreliable.

## Difficulties with Measures of Use

Per capita measures of use, which are intended to facilitate comparisons among libraries, may be inherently biased and misleading. Although it is generally accepted that the extent to which a library is used is an indication of the magnitude of the demand for services that has been met, the performance of the library would be more appropriately measured in terms of the relative proportions of total demand both met and unmet. Still comparable measures of use across libraries are accepted and widely used as indirect measures of library performance. Research (8,9,10), however, has failed to demonstrate that libraries that perform better are used more. The choice by patrons of which library to use appears to be motivated primarily by convenience and ease of access. Furthermore, the extent to which patrons use the selected library appears to be related to the characteristics of the patrons and not to the characteristics of the libraries. Since research has also documented that the proportion of a community which uses the library is a function of the aggregate social, psychological, and educational characteristics of that community, comparative measures of use for libraries from different communities are very problematic. (11,12)

**Registered Borrowers as a Percentage of Population.** This figure, when current, simply measures the segment of the community registered to borrow materials. Given

34

differences in registration policies and procedures among libraries, as well as the mobility of American society, comparisons of registration figures across libraries from different communities is also problematic.

**Annual Library Visits Per Capita.** Annual library visits measures (or estimates) the number of patrons entering the library during the course of one year. This measure, a function of the number of people in the community who use the library and the number of times that each of these patrons actually visit the library, is a recommended criterion for comparisons of performance among libraries. Research, however, has indicated that there are no significant differences among libraries in terms of the mean number of times that patrons visit the library. (9,10) Consequently, differences in annual visits per capita within different communities suggests that what is observed is due to differences in the number using the library. Given that the proportion of residents who use the library is a function of the social, psychological and educational characteristics of the community, the validity of annual visits per capita as a comparative measure of library performance depends on the comparability of the characteristics of the communities served by the libraries.

**Circulation Per Capita.** Perhaps the most widely used measure of library activity and, by unsubstantiated inference, library performance is circulation per capita. Libraries with higher circulations per capita are generally considered performing better than libraries with lower circulations per capita. This inference is dubious. Once again, research has shown no significant differences among the mean number of items borrowed by patrons visiting different libraries within systems. (9,10) Comparisons between systems have yet to be reported. Since there do not appear to be differences among libraries in terms of the mean number of times that patrons report visiting the library and, since there do not appear to be differences among libraries in the mean number of items that patrons borrow per visit, circulation per capita is most likely an indirect measure of the proportion of a service population which uses a library.

**In-Library Materials Use Per Capita.** In-library materials use measures the number of items used, but not borrowed, by library patrons. While this is an important and useful indicator of collection use and staff workload, it does not necessarily measure performance. A potentially more descriptive measure and one that is comparable across libraries, would be the number of items used per visitor rather than per capita.

**Turnover Rate.** Turnover rate--annual circulation/size of collection--is a measure of the frequency with which the average item in the collection circulates each year. The usefulness of this measure as a comparative indicator of performance among libraries is also open to question. Research suggests that differences among library turnover rates may be a function of differences in annual circulation rather than a function of collection size. (13) Given the observations that there do not appear to be any meaningful differences among libraries in terms of the mean number of patron visits, that there do not appear to be differences among libraries in the mean number of items which patrons borrow per visit, and that the proportion of the community which uses the library may be related to social processes beyond the control of the organization, the turnover rate may not be an indicator of library performance.

**Reference Transactions Per Capita.** Notwithstanding the difficulty of collecting accurate tallies of reference questions asked, and notwithstanding the sometimes ambiguous distinctions in coding and counting different kinds of reference questions, the annual number of reference transactions is purported to be a measure of the use of a library as an information retrieval system. Adjusting this tally by the population served allows comparative assessments among different libraries. However, given the discussion above about the proportion of the community which uses the library and the lack of differences among libraries in mean measures of use, differences among libraries cannot be construed as indicating differences in performance among libraries.

**Program Attendance Per Capita.** Program attendance as a measure of library use by patrons is subject to the limitations already noted.

### Difficulties with Measures of System/Service Performance

The title fill rate purports to be a measure of library performance in providing a document on demand. However, the title fill rate, as operationalized in *Output Measures For Public Libraries*, is a patron-reported measure of success in finding a specific title. This is a function both of the library's ability to make the document available (library performance) and the patron's ability to negotiate successfully the catalog and the shelving arrangement in the library (patron performance). Consequently, the fill rate is not a valid measure of library performance. The title fill rate is an unreliable indicator of patron success in finding specific titles; its use in statewide standards and funding formula cannot be justified. (13)

The ability of the library to provide documents on demand, however, is an output of the library as a document retrieval system and is one valid measure of library performance. This output is a result of a set of activities, so it is dependent on a series of evaluations of each of them. As proposed by Ernest R. DeProspo, Ellen Altman and Kenneth E. Beasley (14) and analyzed by F. Wilfrid Lancaster (2,3) and Paul B. Kantor (15), these activities include collection evaluation (the probability that the library owns the item of interest); catalog evaluation (the probability that the item has been correctly cataloged); and shelf availability (the probability that the owned and cataloged item is on the shelf at the time of the request). Using these procedures it would be possible for a library to develop an estimate of the availability of titles sought by patrons and an estimate of the proportions of failed searches that were due to either library failure or patron failure. These data could provide the bases for useful measures of library performance.

For reasons noted above, the subject and author fill rates are also invalid measures of library performance and unreliable measures of patron success within libraries. While it would be possible to evaluate the set of activities that need to be performed to retrieve a subject and author similar to the set of activities for the title fill rate, the methodological problems of intervening in failed subject and author searches and subsequently assessing the relevance of materials provided to the patrons by the staff would probably preclude the development of an easily used measure of library performance. (13)

The browsers' fill rate is an unreliable and arguably trivial measure of library performance.

Document delivery measures the library's ability to provide materials on request within 7, 14, 30, and over 30 days. It is a potentially useful measure in evaluating the various activities related to the library's ability to provide materials on demand. It is germane to the analysis and evaluation of the library as a document retrieval system. However, to the extent that this measure is affected by the response of patrons from whom the material is being recalled or by the response of other libraries from which the document is being borrowed through interlibrary loan, this measure may be confounded.

The reference completion rate--i.e., the number of reference questions answered/the number of reference questions asked--purports to be a measure of the

36

performance of the library as an information retrieval system. Given that this measure does not assess the quality of the information provided in terms of its relevance, accuracy, completeness or timeliness, its usefulness as a measure of library performance is open to question.

### What Can We Conclude Are Output Measures for Public Libraries?

While there is little doubt that measurement of outputs is both managerially necessary, politically useful and socially valuable, the measures proposed and operationalized in *Output Measures for Public Libraries* have serious limitations. Output measures indicate the extent of various library activities which can be useful for justifying the acquisition of additional resources and for scheduling staff. However, the extent of activities experienced by a library appear to be a function of the demand for services by the community, rather than a positive reaction to the quality of services provided by the library. Consequently, comparisons of the measures of use among libraries for the purpose of performance evaluation is a dubious undertaking. In addition, the comparison of per capita measures of use, without taking into account the differences among the communities which are being compared, will further confound the evaluation of performance.

Output measures of library performance are addressed only to a model of the library as a document/information retrieval system. The materials availability fill rates are invalid and unreliable indicators of that performance. Use of the fill rates, as operationalized in *Output Measures for Public Libraries*, for comparative assessments among libraries within systems, for comparative assessments among systems, and in state standards cannot be justified. The measure of document retrieval is a potentially useful indicator of library performance, albeit very much affected by the response times of patrons from whom the materials are being recalled or libraries from which the materials are being borrowed.

### Directions for Research

Directions for research include the development of procedures for comparing levels of library activity across dissimilar communities--for example, regression models where differences in the demographic characteristics of communities can be controlled statistically so that differences among libraries can be identified; the development of valid and reliable measures of comparative performance for the document/information retrieval services of the library; the development of valid and reliable performance measures of the other service programs of the library; and ultimately the development of procedures for measuring the outcomes of library services and programs.

### Beyond the Document Retrieval Model of the Public Library

In 1983, Martin articulated the traditional missions of the public library in society. (16) The Public Library Development Project formalized these missions into a set of eight roles that the library could use for planning purposes. (17) These roles include the library performing as a community activities center, a community information center, a formal education support center, an independent learning center, a popular materials library, a preschoolers' door to learning, a reference library, and a research center. While the model of the library as a document retrieval system underlies all of these, roles help to describe the needs of the community which the library is trying to address. They provide a useful context for planning and evaluation.

By selecting which roles to perform the library is better able to focus its mission in the community, to communicate this mission, to plan activities to produce the desired outcomes for each mission, to determine what resources are needed to perform required activities, and to allocate resources to reflect the relative emphases of the selected missions. In effect, the missions provide the bases for identifying the intended outcomes of library services, for developing performance measures for these outcomes, and for extending the evaluation process to include some indicators of impact on the community or selected segments of the community.

There are three steps to the successful use of roles for planning and evaluation purposes. First, the library needs to determine which roles are most congruent with the community's needs. Second, the library needs to determine which services support which roles and then allocate resources to provide services in support of them. Third, the library needs to evaluate not only how well each service is performing in support of each role but also the impact of each role on the community.

### A Case Study in the Use of the Roles

Eleanor Jo Rodger, Executive Director of ALA's Public Library Association, John Bryson, Director of Planning at the Hubert H. Humphrey Institute of Public Affairs, University of Minnesota, and I recently assisted the Saint Paul (MN) Public Library (SPPL) in developing a new strategic plan. (18) Over the past several years, the library has experienced a considerable increase in demand for services while operating with a no growth budget. The library's intent was to position the library to meet this demand. In order to gather the data to support the planning process, the library commissioned a survey of patrons to determine which roles, current services, and programs were most important to them and how well each service was evaluated by the patrons. Under current conditions, the library was not in a position to survey current nonusers to identify potentially unmet needs in the community, nor did the library attempt to identify new demands for services from among current patrons.

**The Patron Survey**. As preparation for the survey, interviews were conducted with groups of library patrons to solicit lists of services--materials, staff, programs and facilities--that they currently used. From these lists, 34 different services were identified by the library's planning committee as being central to the missions of the library. These services provided the pool of items used to develop the questionnaire for the systemwide patron survey.

The importance of each of the 34 services was rated by the patron on a five-point scale with response categories ranging from "not important to me" to "extremely important to me." These services were evaluated using a five-point scale with response categories ranging from "poor, it needs a very great deal of improvement" to "excellent, a standard for other services." The patrons were also provided with descriptions of each of the eight library roles and asked to rate their importance using the same five-point scale. These data were analyzed to determine which roles were most important to patrons, which services supported which roles, and how each of these services was evaluated by the patrons.

**The Importance of Each Role to the Patrons of SPPL**. The mean importance scale scores from the entire sample of patrons (n = 1036) for each of the proposed library roles are presented in Table 1.

### TABLE 1

### PATRONS' RATINGS OF THE IMPORTANCE OF LIBRARY ROLES RANK ORDERED BY MEAN IMPORTANCE SCALE SCORES

| | |
|---|---|
| REFERENCE LIBRARY | 3.66 |
| POPULAR MATERIALS LIBRARY | 3.65 |
| INDEPENDENT LEARNING CENTER | 3.47 |
| RESEARCH CENTER | 3.43 |
| FORMAL EDUCATION SUPPORT CENTER | 3.30 |
| PRESCHOOLERS' DOOR TO LEARNING | 3.10 |
| COMMUNITY INFORMATION CENTER | 2.71 |
| COMMUNITY ACTIVITIES CENTER | 2.24 |

Further analyses of the data revealed differences between the responses of patrons from within the main library and the responses of patrons from within the branch libraries, and differences in the responses of various demographic segments of patrons. It was possible not only to fine tune the library's roles by type of service unit but also by the segment of the population served.

**Services Related to Each Role.** The patrons' importance ratings for each service were correlated with the patrons' importance ratings for each role are shown in Table 2.

Correlation coefficients, indicate that, with a few notable exceptions, most of the relationships between services and roles are weak; only 21 of the 272 coefficients are greater than .30 and 26 of the services are correlated at .20 or higher with three roles or more. This suggests that for patrons either the roles are not clearly differentiated or that many services support more than one role.

To sort through these relationships the data were submitted to stepwise multiple regression analyses, a statistical procedure which identified for each role that service which was most strongly related to the role, followed by that service which was next most strongly related, followed by that service which was next most strongly related and so on until the procedure did not identify any other related services. The amount of variation in the dependent variable explained by the independent variables in the regression model is measured by the coefficient of multiple determination, $R^2$.

Results of the regression analyses give preliminary indications of the services most closely related to the eight roles of the public library as obtained from the sample of SPPL patrons.

## TABLE 2

## MATERIALS AND SERVICES RELATED TO LIBRARY ROLES

| | Com Act | Com Inf | Form Ed | Ind Ed | Pop Mat | Pre Door | Ref Lib | Res Lib |
|---|---|---|---|---|---|---|---|---|
| 1. mystery/romance/west | | | | | | | | |
| 2. general fiction | | | | | | .23 | | .25 |
| 3. humanities/soc sci | | .23 | .23 | .26 | | | .32 | .31 |
| 4. business/sci/tech | | .20 | | .25 | | | .25 | .27 |
| 5. periodicals | | .22 | | | | | .25 | .21 |
| 6. reference collection | | .28 | .25 | .29 | | | .42 | .40 |
| 7. current information | | .25 | | .25 | | | .29 | .23 |
| 8. records/audio/discs | .20 | .20 | | | | | | |
| 9. videotapes | .21 | .21 | | | | | | |
| 10. government documents | .20 | .25 | .21 | | | | .22 | .28 |
| 11. adult ed materials | .27 | .22 | .20 | | | | | |
| 12. computer software | .24 | .22 | .22 | .20 | | | | .22 |
| 13. teenagers' books | .24 | | .27 | .20 | | .27 | | |
| 14. childrens' books | .21 | | .24 | .24 | | .55 | .21 | |
| 15. community information | .39 | .41 | .22 | .23 | | .31 | .23 | |
| 16. current events/issues | .27 | .34 | .26 | .24 | | .23 | .25 | .26 |
| 17. telephone reference | | .23 | .20 | .21 | | | .28 | .26 |
| 18. professional librarians | | .20 | .25 | .24 | .22 | | .31 | .32 |
| 19. well-trained staff | | | .24 | .24 | .22 | .22 | .34 | .32 |
| 20. courteous staff | | | .20 | | .23 | | .24 | .24 |
| 21. short lines | | | | | .23 | | | |
| 22. computerized search | | .24 | .25 | .27 | .21 | | .29 | .32 |
| 23. week-end hours | | .23 | .26 | .27 | .23 | | .25 | .26 |
| 24. evening hours | | | .27 | .31 | .27 | .22 | .24 | .27 |
| 25. morning hours | | | | | | | | |
| 26. afternoon hours | | | | | | | | |
| 27. adult programs | | .34 | .32 | | | | | |
| 28. children's programs | .29 | .20 | .24 | .23 | | .52 | | |
| 29. clean buildings | .21 | .22 | .22 | .21 | .23 | .20 | .24 | .20 |
| 30. parking | | | | | .20 | | .20 | .20 |
| 31. work space | .21 | .25 | .23 | .24 | .20 | | .28 | .31 |
| 32. meeting rooms | | .42 | .32 | | | | | .21 |
| 33. comfort/pleasant | .24 | .23 | | .20 | .22 | | .23 | |
| 34. collection arrangement | | | | .21 | .27 | | .25 | .21 |

*Community Activities Center Role.* Services: Meeting rooms, information about community groups and activities, lectures, and adult programs ($R^2$ = .26)

*Community Information Center Role.* Services: Information about community groups and activities, lectures, adult programs, reference collections, meeting rooms, book lists about current events and issues, and weekend hours ($R^2$ = .26)

*Formal Education Support Center Role.* Services: Collections of books for teenagers, weekend hours, reference collections, buildings that are clean and well maintained, humanities and social science collections, children's programs and professional librarians to assistant patrons in finding materials and information ($R^2 = .18$)

*Independent Learning Center Role.* Services: Evening hours, reference collections, children's programs, business/science/technology collections, comfortable furnishings and a pleasant atmosphere in each library, general fiction (literature) collections, computer software collection, and current information sources ($R^2 = .21$)

*Popular Materials Library Role.* Services: Evening hours, an arrangement in which it is easy to find materials, collections of mysteries/romances/westerns, general fiction (literature) collections, collection of books for teenagers, current information sources, and check-out lines that are short and move quickly ($R^2 = .17$)

*Preschoolers' Door to Learning Role.* Services: Children's book collection, children's programs, staff that are knowledgeable and well-trained, comfortable furnishings and a pleasant atmosphere, and general fiction (literature) collections ($R^2 = .36$)

*Reference Library Role.* Services: Reference collections, staff that are knowledgeable and well trained, general fiction (literature) collections, comfortable furnishings and a pleasant atmosphere, humanities and social science collections, children's book collection, current information sources, and computerized data base searching ($R^2 = .28$)

*Research Center Role.* Services: Reference collections, staff that are knowledgeable and well trained, computerized data base searching, humanities and social science collections, adequate work space and seating, and computer software collection ($R^2 = .25$)

**Patrons' Evaluations of Services.** When patrons were asked to evaluate the quality of the same set of services that they had rated for importance, the mean evaluation scale score calculated did not include the responses of patrons who indicated that they rarely, if ever, used the service.

These assessments of services could be used to evaluate the library's performance in either of two ways. First, the sample of patrons could be subdivided into subsamples, representing the patrons from different service units, and mean evaluation scale scores could be calculated for each subsample. By matching the evaluation scores to the importance scores for each service the library could identify those services which were in balance and in no apparent need of attention, those services where performance apparently exceeded importance, and those services where performance was not equal to importance and, therefore, in need of attention. Second, using the profiles of services related to the various roles, the library could evaluate its performance in providing those services which support the roles the library had chosen to play. The extent to which the services supporting each role are performed well becomes a measure of the extent to which the roles are performed well.

## New Directions

The evaluation of public library services is currently driven by a model of the library as a document retrieval system and, in the context of this model, library services are evaluated primarily in terms of outputs. While the preeminence of the document retrieval function of the library is recognized, it would be useful now to expand the scope of the evaluation of library services by including other functions of the library, such as those suggested in *Planning and Role Setting for Public Libraries.* While the measurement and evaluation of outputs are clearly necessary and potentially useful, we must turn our attention to extending the evaluation process beyond output measures to include analyses of outcomes.

### Notes and References

1. Mohr, Lawrence B. *Impact Analyses for Programs Evaluation* (Chicago: Dorsey Press 1988).

2. Lancaster, F. Wilfrid. *The Measurement and Evaluation of Library Services* (Washington: Information Resources Press, 1977).

3. Lancaster, F. Wilfrid. *If You Want to Evaluate Your Library* ... (Champaign: University of Illinois, 1988).

4. Lancaster, Measurement: 2.

5. Lancaster, Measurement: 323.

6. Lancaster, Measurement: 4

7. Van House, Nancy; Mary Jo Lynch; Charles R. McClure; Douglas L. Zweizig; and Eleanor Jo Rodger. *Output Measures for Public Libraries* (Chicago: American Library Association, 1987).

8. D'Elia, George and Sandra Walsh. "User Satisfaction With Library Service - A Measure of Public Library Performance?" *Library Quarterly* 53 (April 1983): 109-133.

9. D'Elia, George, and Sandra Walsh. "Patrons' Uses and Evaluations of Library Services: A Comparison Across Five Public Libraries," *Library and Information Science Research* 7 (January - March 1985): 3-30.

10. D'Elia, George, and Eleanor Jo Rodger. "Comparative Assessment of Patrons' Uses and Evaluations Across Public Libraries Within a System: A Replication," *Library and Information Science Research* 9 (January - March 1987): 5-20.

11. Douglas Zweizig. "Predicting Amount of Library Use: An Empirical Study of the Role of the Public Library in the Life of the Adult Public." Unpublished dissertation, Syracuse University, Syracuse, New York, 1973.

12. D'Elia, George. "The Development and Testing of a Conceptual Model of Public Library User Behavior," *Library Quarterly,* 50 (October 1980): 410-430.

13. D'Elia, George. "Materials Availability Fill Rates: Additional Data Addressing the Question of the Usefulness of the Measures," *Public Libraries* 27 (January/February 1988): 15-23.

14. De Prospo, Ernest R.; Ellen Altman; and Kenneth E. Beasley. *Performance Measures for Public Libraries* (Chicago: American Library Association, 1973).

15. Kantor, Paul B. *Objective Performance Measures for Academic and Research Libraries* (Washington: Association of Research Libraries, 1984).

16. Martin, Lowell. "The Public Library: Middle-age Crisis or Old Age?" *Library Journal* (January 1, 1983): 17-22.

17. McClure, Charles R.; Amy Owen; Douglas L. Zweizig; Mary Jo Lynch; and Nancy A. Van House. *Planning and Role Setting for Public Libraries* (Chicago: American Library Association, 1987).

18. D'Elia, George, Eleanor Jo Rodger and John Bryson. "Saint Paul Public Library Strategic Planning Project, Phase I: Survey and Evaluation." Unpublished report, Minneapolis, Minnesota, 1988.

# IMPROVING STATE LIBRARY EVALUATION OF FEDERAL PROGRAMS

**Charles R. McClure**

## Abstract

This paper explores selected issues and approaches regarding the evaluation of programs and library development. First it identifies recent events that have pressured the state library agencies (SLAs) to improve their evaluation activities. Next, suggestions are offered to clarify the factors affecting the SLAs' evaluation processes. A general model depicting the SLAs' role in evaluation is proposed to identify areas where the process can be enhanced. A number of steps are recommended to improve the quality of evaluations.

In recent years, the state library agencies (SLAs) have assumed responsibility for the development and promotion of a broader range of library services. Yet, paradoxically, the SLAs are one of the least understood and least studied organizations in the library-information services arena. They find themselves in a complex political environment and frequently must resolve issues and make decisions which cannot possibly please everyone, a good example being the development of statewide standards.

A discussion of the range of activities and services for which the SLAs have responsibility--beyond the scope of this paper--is available elsewhere. (1) However, one primary responsibility of the SLAs is to distribute fiscal support for statewide library development. Sources for these funds are each state's aid to libraries and federal monies allocated to the states primarily through the Library Services and Construction Act (LSCA). (2) Of special interest here is the responsibility of the SLAs in evaluating the success with which LSCA monies do, in fact, promote library development.

The amount of direct state aid to public libraries varies considerably--from nothing, to a few hundred thousand dollars, to tens of millions. (3) LSCA funding for library development, primarily through Title I (support for a broad range of public library services, including services to groups with special needs), Title II (construction), and Title III (programs supporting multitype library cooperation and resource sharing), is a significant amount. The total LSCA authorization requested in the 1990 budget is $137.2 million. (4)

Many librarians argue that state aid and LSCA funds are inadequate to meet the needs for successful library development and additional funding is necessary. To support that argument the SLAs must offer specific evidence, especially quantitative evidence, to demonstrate that LSCA has a significant impact on improving statewide library services.

Indeed, there is some irony in the need for LSCA evaluation data. Not only must the SLAs prove that there is improvement resulting from federal funds, they must simultaneously demonstrate the ongoing need for more funds to fuel continued improvement. As one ex-state library official commented, "We need at least two evaluation measures: One to demonstrate the success from funding and another to prove the continual inadequacy of library services." Regardless of the type of evaluation data needed, a 1985 study of LSCA and SLAs, however, concluded that:

> Program evaluation has certainly been the weakest part of the SLA's activity regarding the implementation of LSCA Title I.... Considering program evaluation to be impossible, SLAs have done little of it. The skills needed for program evaluation are lacking in the SLAs, and no SLA has established or developed an evaluation unit. (5)

Although this situation may have improved since 1985, the efforts by many SLAs to demonstrate the adequacy or inadequacy of state aid or LSCA funding to state and federal policymakers have been weak, at best.

The evaluation process for federally funded projects can be improved; greater attention must be given to implementing program assessments. It is essential to demonstrate the success and impact of state aid and to provide data by which library development can be enhanced. Before improvement occurs, however, both state and federal officials must better cooperate in the development of evaluation techniques.

## Key Factors Affecting SLA Evaluation Activities

A number of key factors can be identified that have brought increased pressure on the SLAs to better evaluate the success of state and federal aid. They may have different impacts on and importance for a particular SLA, but, when taken together, they describe issues and concerns that are shaping the SLA's role in dealing with evaluation responsibilities.

**Federal and State Appropriations.** During the 1980s, the library community found it increasingly difficult to maintain or increase state and federal aid to libraries. The Reagan administration sent clear signals to the SLAs that funding for LSCA and other federal library programs should be reduced if not eliminated. Annual legislative fights began and treks to Washington by various library leaders trying to put LSCA and other library programs back into the federal budget became commonplace. Although funds to support library programs were returned to the budget by the Congress, the evidence needed to demonstrate specific impacts and benefits from these programs in general, and LSCA in particular, was scarce and largely anecdotal.

In 1988 the federal government proposed to replace LSCA programs with the new Library Services Improvement Act (S. 2579). The American Library Association (ALA) and the Chief Officers of State Library Agencies (COSLA) opposed the proposal. (6) The implications and impact of the replacement legislation for LSCA were unclear. The point, however, is that there is an ongoing political process in which the library community must defend existing aid for libraries, demonstrate the importance and impact of such aid, and in recent years, fight to maintain the status quo as opposed to obtaining additional or new sources of federal support.

At the state level, the 1980s have produced a mixed bag for support to libraries. Depending on the state in question, there are examples of stagnant or reduced appropriations as well as increasing appropriations. Generally, however, state appropriations to library development have not kept pace with the need for library services. And, similar to the case at the federal level, there is an ongoing political process in which the state library community defends existing aid, attempts to demonstrate its impact and importance, and justifies the need for additional appropriations.

**Accountability.** In 1983, the Inspector General for Audit at the U. S. Department of Education issued a report that reviewed the Illinois SLA's administration of the LSCA program. That report concluded federal guidelines for distributing money had not been followed, Title I funds had been used for projects not meeting the legislative intent of LSCA, and more attention was needed to ensure that funded projects met the needs of targeted or special groups. (7)

The report triggered an immediate debate and some degree of controversy regarding the role of the SLAs in administering LSCA money. Many state library officials were surprised that such an audit occurred and that a report was issued at all. In effect, the SLAs were put on notice that they were responsible for demonstrating accountability for the large amounts of LSCA funds being distributed throughout their states. This message did not go unnoticed by SLAs outside Illinois.

At the state level, informal conversations with a number of SLA directors suggest that there is increased pressure to demonstrate accountability for the effective spending of state aid. These pressures come from stricter accounting practices, statutory requirements to demonstrate effectiveness of state-sponsored programs and, in many states, increased competition for public monies as well as demands that the funds help resolve a broad range of social problems.

The accountability pressure also is felt within each state's library community where different stakeholders request funding to support a range of different activities. Since the SLAs cannot satisfy all those demands, some projects go unsupported. As a result, the SLAs must be able to demonstrate the effectiveness and impact of programs that are supported. If, in fact, the programs supported are not effective and have little impact, the library community is likely to apply pressure on the SLAs to redirect their funding priorities.

**Evaluation Skills.** In the library profession as a whole there has been a tradition of limited interest in ongoing evaluation research and the rigorous assessment of library services and programs. (8) As a result, there is limited knowledge and understanding of the evaluation process and specific evaluation techniques. The SLAs typically have few staff with adequate training to develop appropriate evaluation designs, implement the evaluation process, and analyze the results of those evaluations.

But even if the SLAs can overcome the problem of limited evaluation skills and knowledge within their own agencies they have an additional hurdle to overcome--the SLAs are often dependent on the libraries that receive the aid to conduct their own evaluations. This dependency frequently injures further the effectiveness with which the evaluation is done because, similarly, there are few individuals in local libraries trained in the evaluation process or able to implement an effective evaluation.

**The Public Library Development Project.** As a result of an ongoing effort, orchestrated by the Public Library Association (PLA) the American Library Association (ALA) published two manuals in 1987, *Planning and Role Setting for Public Libraries: A Manual of Options and Procedures* (9) and *Output Measures for Public Libraries: A Manual of Standardized Procedures*, second edition. (10) More than 10,000 copies of each manual have been sold, numerous workshops describing their use have been conducted, and increased attention is being given to public library planning and use of output measures.

A number of SLAs have promoted the manuals as a means of encouraging statewide public library development. Some like Colorado, have designed statewide data collection activities that include output measures and are producing annual reports describing libraries in terms of output measures. (11)

The manuals, however, were developed primarily for individual library use and not as a vehicle for SLAs to promote statewide measurement of public library services. Further, the production of output measures constitutes but one possible technique to conduct an evaluation. On one hand, there is increased awareness of planning and use of output measures. On the other hand, many SLAs are unsure: 1) how, if at all, they should use these manuals for statewide library development; 2) if they should incorporate output measures as standards; and 3) how output measures might assist them in evaluating library services, programs, and overall development.

**Other Factors.** Clearly, other factors such as inadequate SLAs' staffing given the SLAs' range of responsibilities, the increasing paperwork requirements to be met by the SLAs in administering both state and federal aid, limited training opportunities from library schools and other sources in the area of evaluation, and the minimal interest/capability on the part of local libraries to spend scarce resources on program assessment are also important to evaluation activities. Yet given these recent developments, the issue to be resolved remains: How can federal officials and the SLAs better evaluate the success of state and federal aid for library development?

### Basic Issues Affecting Evaluation Activities

The degree to which the SLAs identify and resolve the key issues in evaluation will determine the overall effectiveness with which the evaluation process is designed and implemented. To improve the process, it is important that the various players involved are agreed on the meaning and purposes of evaluation. Evaluation is a systematic process that assesses a particular service, activity, or program in terms of certain criteria and offers a judgment of the value of that service, activity, or program. (12) Minimally, evaluation includes the following activities:

- Identifying and collecting data that describe specific services, activities, or programs;
- Establishing criteria by which the services, activities, or programs can be assessed; and
- Making judgments about the degree to which the data indicate that a service, activity, or program meets the criteria.

In one sense, the evaluation process primarily serves as an information gathering, analysis, and reporting function. Sprinkled throughout this seemingly rational and objective process are individual value preferences and political judgments. The SLAs and federal officials are also confronted with a range of possible purposes for conducting evaluations:

- Satisfying state and federal legal requirements as a condition of obtaining the aid;
- Proving the quality or success of particular library services, activities, and programs;
- Demonstrating accountability to outside agencies;
- Justifying decisions made in the allocation of state and federal resources;
- Requesting additional resources;
- Defining and identifying library development needs; and
- Ensuring that libraries met state guidelines or standards.

Other participants in the evaluation process may have other purposes. Further, participants may have different long range agendas for developing library services. So it is essential that the SLAs and the federal government clearly define what they mean by the term evaluation and explain their purposes for conducting a specific evaluation activity.

**Level of the Evaluation.** Another point that the SLAs and federal officials must consider is the level of the evaluation. Minimally, there are at least three:

- Evaluation of overall federal/state appropriation programs, e.g., Title I of LSCA or a particular state program such as records/archives management;
- Evaluation of local library programs, awards, and services, e.g., a literacy grant awarded to a particular library by either a SLA or a federal agency;
- Evaluation of SLA's operations, e.g., evaluation of how well a SLA administered a specific program, such as implementation of statewide standards or administering LSCA funds.

The SLAs and federal officials must clarify the level of a particular evaluation before they can design the most appropriate evaluation.

**Types of Evaluation Measures.** SLA staff, state and federal officials, and librarians must also consider the types of evaluative measures that can be used to assess a library service, program, or activity. Instead of merely collecting data, the focus should be on first, determining which type of measure is needed, then obtaining the necessary data. One typology for categorizing types of measures includes:

- Extensiveness measures: Focus attention on "how many" occurrences of a service or activity were provided within a specified time period.
  EXAMPLE: Number of reference transactions per day.
- Efficiency measures: Focus attention on the amount of resources or the use of resources for a particular service or activity.
  EXAMPLE: Cost per program attendee.
- Effectiveness measures: Focus attention on how well the service did what it was intended to do, usually in the context of accomplishing stated service, program, or activity objectives.
  EXAMPLE: Percentage of reference questions answered correctly.
- Impact measures: Focus on the benefit or result of a service, program, or activity; impacts usually cannot be measured until a suitable period of time has passed following the implementation of the service, activity, or program.
  EXAMPLE: Percentage of librarians attending a workshop on how to successfully weed a collection that did, in fact, implement a weeding program within six months of the workshop.

In reality, the lines that separate one type of measure from another are not always distinct and some measures may, in fact, address two of these categories.

At this point in time, state and federal officials need to give greater attention to understanding the differences among these types of measures and to conducting library evaluations that go beyond extensiveness measures and move into efficiency, effectiveness, and impact. Moreover, efficiency, effectiveness, and especially impact measures are much more likely to convince other individuals of the relative success and value of a particular library service, program, or activity. SLAs need to clarify which measures are needed for what types of evaluations, both for themselves, and for the libraries whose evaluation efforts they guide.

**Responsibility for the Evaluation.** Another key issue is determining responsibility for conducting the evaluation. The following stakeholders are conceivable:

- Federal government officials
- State government officials
- SLA staff
- Library staff
- Consultants
- Local library community members or governing boards.

Until responsibilities for specific evaluation activities are clarified among these and possible other stakeholder groups, evaluating the success and value of LSCA funds will remain difficult.

**Use of Output Measures.** Output measures are a type of performance measure that focuses attention on assessing the results or products of a particular library. They are intended to be an objective source of management information that can assist library decision makers to improve the performance of their libraries.

A number of SLAs are considering the use of output measures for statewide assessment of libraries and library services. The states of Illinois, Oklahoma, Utah, Wisconsin, Colorado, and Ohio (13) have produced written manuals or reports that collect these measures for planning and implementing statewide standards. According to a 1988 study, 18 states now require public libraries to submit information for at least one output measure. (14) In a number of these instances, output measures are linked to statewide standards.

The use of output measures by SLAs for statewide as opposed to individual evaluation of libraries and library services is fraught with problems. A discussion of some of these problems appears elsewhere. (15) However, major concerns include:

- Original purpose of output measures. The primary intent in producing output measures was to supply a tool that would support planning and evaluation in an individual library setting, i.e., to serve as self-diagnostics for the library to compare its performance against itself over time.
- Noncomparability among output measures across libraries. Because libraries may collect, code, and report data to compute output measures differently, it is unlikely that the measures can be compared meaningfully one to another--even for libraries in like demographic situations; this is an especially important consideration for SLAs that want to develop statewide standards based on output measures.
- A partial and incomplete assessment of performance. Output measures provide only one type of indicator for assessing a library or library services; meaningful assessment of a particular library or library service requires the use of a range of indicators with multiple measures over a period of time.

- Link between output measures and library mission, roles, and goals. Scores on output measures lose meaning when they are not also attached to the library's mission, roles, and goals. As an example, it may be appropriate for a library to have a low score on collection turnover if library decisionmakers have decided to emphasize the role of community activities center or goals related to programming.
- Confusion over what an output measure actually measures. A broad range of variables affect an output measure; these variables are likely to differ significantly from library setting to library setting. Making comparisons statewide on library output measures is inappropriate unless a range of additional factors such as those shown in Figure 1 are also known.

While output measures certainly are appropriate for gauging the success of a specific program, service, or activity, within a particular library, numerous problems exist in using them to make comparisons nationally or statewide. SLAs should look at broader evaluation issues, methods, and approaches rather than focusing on a particular measurement device, such as output measures, simply because a manual exists for collecting library output measures.

Depending on the purpose and definitions of the evaluation process, the object of the evaluation, the type of measures needed, and the appropriateness of using output measures, the SLAs and federal officials have a range of options for designing the actual evaluation process. It is unlikely that the SLAs can conduct the range and depth of evaluations ideally necessary for all the various programs and services in which they are involved, but in the final analysis it is the SLAs that will most likely choose which evaluation efforts will and will not be conducted.

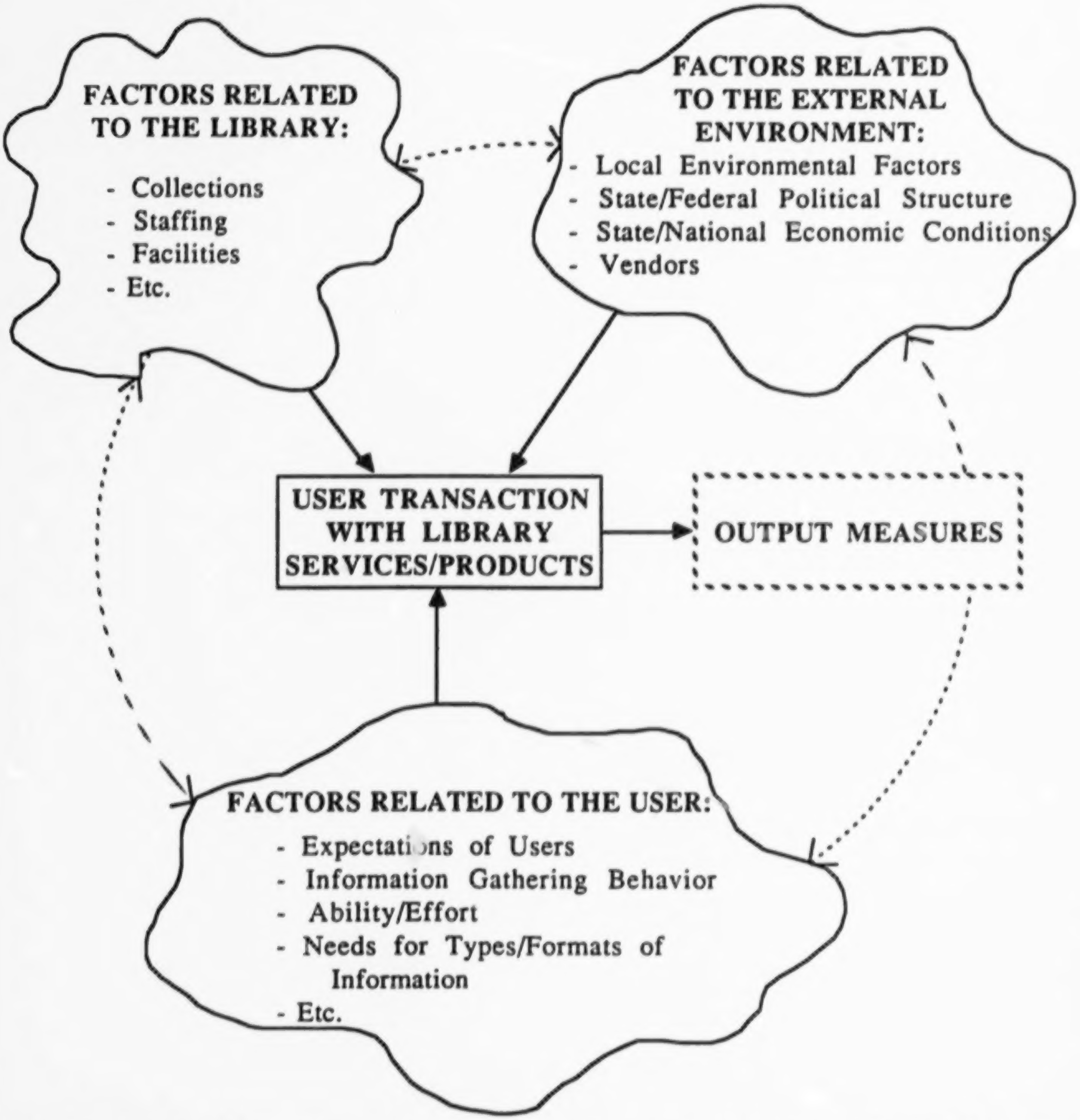
### A General Model of the Evaluation Process

The recent developments affecting evaluation activities and the need to address the major issues described should also be considered in light of the process by which evaluation activities occur. While each SLA may have some unique operating characteristics and each may deal with situations unique to their particular state, there appears to be a general model that describes the federal/state aid planning/evaluation process. Based largely on the author's review of SLAs' documents and federal forms and guidelines (16), Figure 2 illustrates how for example, the LSCA evaluation process operates.

Differences occur from state to state. The intent of this figure is to 1) describe the major activities related to the evaluation activities of the SLAs; 2) identify critical points in the process where intervention strategies might be appropriate to improve evaluation activities; and 3) provide a basis from which the process might be redesigned.

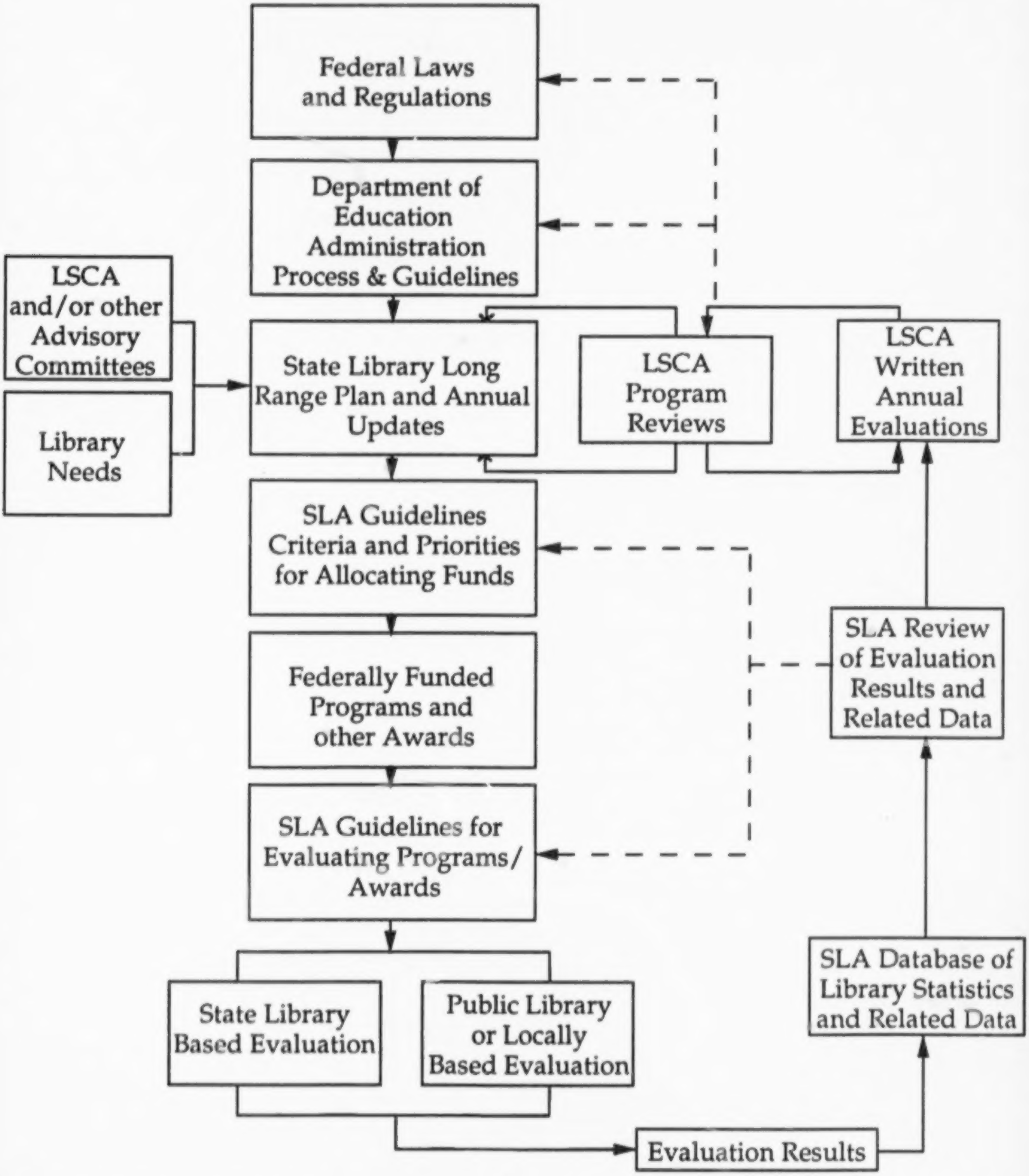
**Description of the Model.** The model, recognizes that the SLAs' responsibilities originate in statutory law or regulations and guidelines from both the state and the federal government. However, these laws and regulations are interpreted and administered by oversight agencies, which in the case of the federal government is the U.S. Department of Education, Office of Library Programs. Under LSCA, for example, each SLA must develop a written long-range plan and submit annual updates to the Office of Library Programs. The SLAs may obtain input in the development of these

**Figure 1**
**User Perspective Toward Library Performance Measurement***


FACTORS RELATED TO THE LIBRARY:
- Collections
- Staffing
- Facilities
- Etc.
FACTORS RELATED TO THE EXTERNAL ENVIRONMENT:
- Local Environmental Factors
- State/Federal Political Structure
- State/National Economic Conditions
- Vendors
USER TRANSACTION WITH LIBRARY SERVICES/PRODUCTS
OUTPUT MEASURES
FACTORS RELATED TO THE USER:
- Expectations of Users
- Information Gathering Behavior
- Ability/Effort
- Needs for Types/Formats of Information
- Etc.

*SOURCE: Nancy A. Van House, Beth Weil, and Charles R. McClure. Output Measures for Academic and Research Libraries (Chicago: American Library Association), in preparation.

**Figure 2**
**Federal Aid Planning/Evaluation Process**


Federal Laws and Regulations
Department of Education Administration Process & Guidelines
LSCA and/or other Advisory Committees
Library Needs
State Library Long Range Plan and Annual Updates
LSCA Program Reviews
LSCA Written Annual Evaluations
SLA Guidelines Criteria and Priorities for Allocating Funds
SLA Review of Evaluation Results and Related Data
Federally Funded Programs and other Awards
SLA Guidelines for Evaluating Programs/Awards
State Library Based Evaluation
Public Library or Locally Based Evaluation
SLA Database of Library Statistics and Related Data
Evaluation Results

plans from advisory committees or from available information describing existing library needs.

Based on the plans, the SLAs typically develop guidelines and priorities for allocating funds. Some of these allocations are mandated at the state level due to an existing formula which apportions monies based on library service area populations, ability to meet statewide standards, or other criteria. The SLAs eventually award or allocate these funds; the SLAs may or may not have specific guidelines to assist grantees in evaluating the success of the activity for which they received fiscal support. After the allocation of monies an evaluation process is prescribed, which the SLAs or the recipient of the award may conduct. In fact, the evaluation frequently has serious flaws in method, or, may not actually be done at all. The evaluation produces results--usually in the form of some data which are managed and organized in a database of statewide evaluation results. This database may take the form of manila folders in a filing cabinet or a machine-readable PC-based spreadsheet or database.

The SLAs can review evaluation results as a basis for revising funding and evaluation guidelines as well for providing input to the LSCA mandated annual written evaluation. LSCA program officers from the Office of Library Programs also review the evaluation results. Their review can provide feedback to revise existing LSCA and state laws and regulations, the administrative process for awarding monies to the SLA, and update long range and annual plans.

**Critical Factors for Improving the Process.** The process suggests that the degree of success likely to result from the evaluation process, minimally, is linked to the success with which certain activities occur. Improvements can begin in places which lend themselves to relatively quick intervention.

1. Clarify State and Federal Intent

The statutory and administrative intent for state and federal aid is oftentimes unclear and interpreted differently by state and federal agencies and program officers. The intent may be in conflict with other responsibilities of the SLAs. As an example, LSCA regulations (detailed in 34 *CFR* 770.21 - 770.23) regarding accountability, periodic evaluation, and reporting are vague and need reworking.

Further, the stated intent may differ from the implementation. In recent years, while U.S. Department of Education spokespersons have talked a great deal about the importance of library program evaluation and the need to perform them better, the evaluation form to assess proposals for the 1989 LSCA Field Initiated Studies program allocated only 5 of 100 possible points for the quality of the proposal's evaluation design. (17) Intent and implementation have to be better coordinated.

Even though the more specific the intent of these laws and administrative guidelines, the less flexibility SLAs might have in implementing them, in the area of evaluation, there clearly is room for both the SLAs and the Office of Library Program to work cooperatively to clarify intent, propose procedures and guidelines, and better articulate their expectations to each other and to the broader library community.

2. Link Evaluation Designs and Measures to the State Plan

The LSCA mandated long-range and annual plans are frequently laundry lists of goals, objectives, and proposed activities. Inclusion of an evaluation design, measures, and procedures for how the objectives will be evaluated should be required at this stage of the process. Developing fewer goals and objectives with more attention to how they will be assessed, early on, in the long-range and annual plan, could improve the evaluation process.

Indeed, it may be a misnomer to refer to the LSCA mandated long-range and annual documents as plans. Typically, they are political statements describing how the SLAs will address the federally-determined priority funding areas for any particular year. Addressing federally-determined priority areas to obtain grants is not the same as having a long-range plan for statewide library development. SLAs should consider separating these two plans both physically and intellectually.

3. Develop Evaluation Training Programs and Written Guidelines

Concern exists at two different levels here. First, staff of the SLAs may need additional training and education regarding evaluation designs, measures, and procedures. The Office of Library Programs can assist the SLAs by providing resources for the development of a straightforward training manual describing a step-by-step approach for evaluating funded programs and activities.

Second, applicants and recipients of state and federal library aid may also lack the necessary knowledge of evaluation designs, measures, and procedures. A requirement that the recipient of an award conduct an evaluation has little impact if the recipient has no knowledge of how to do so. A training manual coupled with carefully developed workshops explaining how recipients can conduct such an evaluation is also needed.

4. Expand the Time-Line for Conducting Evaluation

The current evaluation process is intended to occur over a one-year period. For many SLA award recipients it is extremely difficult to produce meaningful evaluation results within this time frame. Further, evaluation of program effectiveness and impact may not be possible until one or two years after receipt of funds. Federal administrators and SLA staff should consider elongating the time period in which funding recipients can conduct evaluations and report findings.

5. Manage Evaluation Data

Analyzing and interpreting evaluation data requires data management skills and the development of management information systems. Currently, few SLAs have carefully developed programs for collecting, organizing, analyzing, and reporting evaluation data. Neither is it clear how the various federal program officers manage evaluation data. Such a program is essential for:

- SLA staff, federal officials, and individual librarians in the states to have a common base of knowledge to make judgments regarding the success of state and federal aid;
- Librarians to obtain from the SLAs summary reports and customized analyses to assist them in improving their evaluation and improving library services and activities; and
- SLA staff and federal program officers to maintain a corporate memory of evaluation results, relate those results to other library data, and report them as a means to refine and improve the evaluation process.

The SLAs and the U.S. Department of Education need to commit federal and/or state resources for managing the data that arise from the evaluation process. A current move to strengthen LSCA by not only mandating evaluation but also mandating federal funding to support evaluation is a step in the right direction.

### 6. Learn from the Evaluations

Many involved in the evaluation process perceive it as a purely academic exercise. In their view, the SLAs must meet certain requirements--some of which include evaluation--in order to receive state and federal aid. Local libraries must meet certain requirements some of which include evaluation in order to receive state and federal aid. Unfortunately, the emphasis is on evaluation to obtain funds rather than evaluation to improve services or to learn about how to do it better next time.

To learn from evaluations, feedback loops must take on increased importance. Mechanisms such as management information systems must be in place for the four major stakeholders in the evaluation process--state and federal officials, SLA staff, and local librarians--to obtain meaningful reports from the evaluation process, to discuss evaluation results with each other, and to act upon the findings as a basis for future improvements.

## Resolving Issues and Implementing Changes

Clearly, staff of the SLAs, federal program officials, state agency officials, local librarians, and library associations need to work together to address and resolve the issues surrounding the evaluation of federally funded library programs. Some steps can be taken immediately to improve the evaluation process, but generally, a comprehensive program of activities needs to be developed, reviewed, and implemented as opposed to applying band-aid solutions.

Initially, an empirical approach must assess the existing evaluation process. Figure 2 represents a deductive model; data are needed to refine and validate it. Once the actual evaluation process is modeled, it can be carefully reviewed and assessed to find strategies to improve and simplify it. However, it is essential that policy making and research by testimony and opinion do not replace policy analysis and evaluation research by empirical data.

Stakeholders interested in improving the evaluation process can also begin by creatively brainstorming criteria that would describe an ideal process. For example, it might be proposed that the evaluation process should:

- Require as little staff time as possible to implement;
- Be based on procedures that are as simple and straight-forward as possible with paperwork burden of these procedures minimized for all concerned;
- Ensure ongoing and meaningful communication mechanisms among the various stakeholders in the evaluation process;
- Produce data that are useful to all stakeholders for improving library development at the national, state, and local levels;
- Provide for ongoing evaluation education and skills-building for individuals designing, conducting, reporting, or interpreting the evaluation;
- Include a regular review of the evaluation process itself.

While these criteria are illustrative only, in the design of any process, it is necessary to establish requirements that the evaluation process should meet. Now it is unclear what the requirements are for the existing process or the degree to which those requirements are being met. It is clear however, that there is widespread agreement that the existing process for evaluating state and federal funding of library development is ineffectual and riddled with problems. It is also clear that there are a number of possibilities for improving the process. Less clear, however, is the level of commitment and resources stakeholders are willing to dedicate to improving the existing situation. But such commitment is essential, it is needed now, and it is needed from across the range of stakeholders involved in the evaluation process. The library community must demonstrate the success and show the value of federal funding, strive to obtain maximum benefit from these scarce resources, obtain ongoing information to improve the programs and activities funded by state and federal monies, and explain clearly the importance and need for such funds. Improving the process for evaluating the use of state and federal funds is an essential step in accomplishing these objectives.

## Notes And References

1. McCrossan, John A., Editor. "State Library Development Agencies," *Library Trends*, 27 (Fall, 1978), and Charles R. McClure, Editor. *State Library Services and Issues: Facing Future Challenges* (Norwood, NJ: Ablex, 1986) are two sources.

2. 20 *USC* 351 et seq. For a detailed discussion of the evolution of LSCA see Holley, Edward G. and Robert F. Schremser. *The Library Services and Construction Act: An Historical Overview from the Viewpoint of Major Participants* (Greenwich, CT: JAI Press, 1983).

3. Asp, William G. "The State of State Library Agencies." In: Charles R. McClure, Editor. *State Library Services and Issues: Facing Future Challenges* (Norwood, NJ:

Ablex Publishing Corporation, 1986): 41-43. Provides recent overview of state aid expenditures.

4. "Appropriations, FY 1990," *Washington Newsletter*, 41 (March 27, 1989): 1-2.

5. Shavitt, David. *Federal Aid and State Library Agencies: Federal Policy Implementation* (Westport, CT: Greenwood Press, 1985): 124-125.

6. "Appropriations, FY 1990."

7. U.S. Department of Education, Office of the Inspector General for Audit. *Report on Audit of the Implementation and Administration of Library Services under the Library Services and Construction Act for the Period July 1, 1977 through December 31, 1980: State of Illinois, Illinois State Library, Springfield Illinois* (Washington, D.C.: U.S. Department of Education, 1983).

8. McClure, Charles R. and Ann Bishop. "The Status of Research in Library/Information Science: Guarded Optimism," *College and Research Libraries*, 50 (March-April, 1989): 127-143. Gives an overview of the role and status of research in the profession.

9. McClure, Charles R; Amy Owen; Douglas Zweizig; Mary Jo Lynch; and Nancy A. Van House. *Planning and Role Setting for Public Libraries: A Manual of Options and Procedures* (Chicago: American Library Association, 1987).

10. Van House, Nancy A.; Mary Jo Lynch; Charles R. McClure; Douglas L. Zweizig; and Eleanor Jo Rodger. *Output Measures for Public Libraries: A Manual of Standardized Procedures*, second edition (Chicago: American Library Association, 1987).

11. Lance, Keith Curry and Maureen Crocker. *Statistics & Input-Output Measures for Colorado Public Libraries* (Denver: Colorado State Library, 1988).

12. Rossi, Peter H. and Howard E. Freeman. *Evaluation: A Systematic Approach*, third edition (Beverly Hills: Sage 1985). Represents one of a number of good introductory evaluation texts.

13. *The Upgrade Process: Planning, Evaluating, and Measuring for Excellence in Public Library Service* (Salt Lake City: Utah State Library, 1987); *Wisconsin Public Library Standards* (Madison: Wisconsin Department of Public Instruction, 1987); and Lance, Keith Curry and Maureen Crocker. *Statistics & Input-Output Measures for Colorado Public Libraries* are examples of such publications.

14. Weech, Terry L. "Planning Process & Output Measures in State Standards for Public Libraries: Preliminary Report." Presentation at the American Library Association, Annual Conference, PLA Planning Process Discussion Group, January 7, 1989.

15. McClure, Charles R., "The Role of the State Library Agency for Developing Statewide Performance Measures, Planning, and Standards." In: Charles R. McClure, Editor. *State Library Services and Issues: Facing Future Challenges* (Norwood, NJ: Ablex Publishing, 1986): 187-211.

16. These documents include long range plans, annual plans, and annual evaluation forms submitted by a number of SLAs to the U.S. Department of Education, Office of Library Programs and various memos and guidelines distributed by the U.S. Department of Education to SLAs, including "Performance Report for Titles I and III," "Performance Report for Title II," "Financial Report for Titles I and III," and "Long Range Program Review Form PL 98-480."

17. "Field Initiated Studies Program 1989 Evaluation Form" (Washington, D.C.: U.S. Department of Education, Office for Educational Research and Improvement, Library Programs, 1988), 3 page mimeograph.

# EVALUATING THE IMPACT OF FEDERALLY FUNDED PUBLIC LIBRARY YOUTH PROGRAMS

Mary K. Chelton

## Abstract

Youth services programs, primarily in public libraries, are explored in terms of their evaluability with suggestions for federal funds. Despite generally agreed upon long-term impacts, the lack of operational client definitions, lack of technical skill in evaluation, low organizational status, and absent documentation hinder evaluation efforts. Intermediate steps to improve evaluability such as identifying shared uses for an evaluation, improved recognition and publicity, state and/or federally mandated service definitions, and subcontractor academic evaluators are introduced.

Impacts are variably defined in evaluation literature, and by those attempting to achieve them. Here the term means the knowledge, skills, and abilities a young person attains as a result of exposure to, and participation in, a library program. Yet, the intended impacts of library youth services, regardless of setting, are well understood by most competent youth services practitioners. In fact, they are so well understood that, until recently, they were remarkably poorly articulated in professional materials, all of which emphasized how to practice rather than why. There was almost a conspiratorial assumption of shared philosophical understanding with the readers of such works, so that questions about assumptions, almost always coming from a non-youth services source, were seen automatically and defensively as an attack. The fact that these assumed impacts had an almost nonexistent research base in library research literature was of little concern.

The communally self-isolated and self-satisfied youth services librarians knew that what they were doing was right, if only their administrator/principal would give them adequate autonomy and resources. They were adamant in demanding prescriptive standards from outside professional sources which would reinforce the necessity for the allocation of resources they felt were optimal, with little concern for the impact of those allocations; the impacts were taken for granted.

While these attitudes are breaking down somewhat, there are residual effects for impact evaluators to contend with, not the least of which is territoriality. Youth services practitioners feel that they know what works best for kids in terms of library services and they do not appreciate outsiders telling them what they should be doing. Outsiders frequently are subjected to a do-they-like-kids litmus test to achieve credibility before their opinions or advice are heard. Youth services librarians particularly resent the exclusive use of empirical quantification and feel that the importance of what they do is

---

Acknowledgements: The author gratefully acknowledges the assistance of Dr. Wolfgang Pindur, Department of Urban Studies, Old Dominion University, Norfolk, Virginia, in the conceptualization of this paper.

diminished by statistical documentation without at least equal attention to qualitative factors and reporting. They are still frequently very naive about the political processes of their institutions, unable to master how to communicate equally well with their colleagues and administrators as with those to whom they provide direct service--children, adolescents, and their parents, teachers, caregivers, and youthworkers.

Besides these generalized attitudes and characteristics which are not always, but may be, present in any particular evaluation situation, the other pervasive problems impact evaluators must contend with are: 1) The absence of a national operational definition of child and adolescent among youth service librarians which greatly inhibits comparative program analysis; 2) The existence of multiple influences on the intended library impacts such as parents, friends, teachers, socioeconomic status, ability to read English, out-of-school group activities, etc., beyond those supplied uniquely by the library; 3) The multiple manifestations of the intended impacts in areas other than the library, such as in the home or classroom; 4) The time at or during which the intended impacts are expected to become apparent to someone else; and 5) The fact that youth serving librarians in schools and public libraries do many of the same activities to achieve different impacts because of the generic nature of libraries and because of the developmental needs of a shared youth clientele.

## Intended Impacts

Youth services impacts are conceptually divided into three categories which are developmentally based on the nature of the service clientele, regardless of whether they are delivered in a school or in a public library setting. The long-term impacts youth services librarians hope to achieve are:

### Preschool

1) The development of receptive language, i.e., becoming accustomed to the sound and organization of written language prior to the development of actual reading skills.
2) The establishment of routines and habits conducive to reading and learning.
3) The development of a sense of story, i.e., the fact that pictures have meaning; that pictures are different from words; that pages turn from right to left; that books have a front and a back and a right side up; that stories have a structure.
4) The association of pleasure with reading activities. (1,2,3,4)

### Elementary School Age (K--6th grade)

1) Sustained and expanded reading competence skills.
2) Increased reading comprehension.
3) Sustained motivation to read and pleasure in reading.
4) More effective formal instruction in terms of both what is learned and how learning occurs.
5) Maintained and/or increased self-esteem.
6) The acquisition of rudimentary independent research skills.

7) The fostering of creativity through free, independent, and self-directed inquiry. (5,6,7)

Secondary School Age (7th--12th grades)

1) Knowledge of oneself as a unique person with special abilities, who is also part of a local, national, and world community.
2) Reading for pleasure as a lifelong habit.
3) The ability to do independent research using a wide variety of media.
4) The acquisition of knowledge and skills necessary for gainful employment and personal relationships.
5) The ability to distinguish between factually-derived vs. unsubstantiated and/or biased information.
6) The knowledge and skills necessary to participate as a citizen of the United States.
7) The knowledge that, while one cannot know everything, there are places and people to help someone find information and to help promote lifelong learning. (4,8,9,10,11,12,13,14)

The best youth services programs will attempt all of these impacts regardless of setting, but the scope of the school media services program generally is more narrowly focused on instructional design, actual instruction, and the research skills needed by students to acquire a specific body of knowledge in the school setting. The public library program is generally broader in the scope of information, if not formats, available. While the public library provides homework and independent study support for school-age children and adolescents, it is more focused on the frequently unspecified personal interest of young users at home and in the community. Since the activities in the two different institutions can greatly resemble each other, however, another distinguishing conceptual difference is the fact that using the personal interests of young users is a means to an end in the school library; whereas enriching personal interests can be an end in itself in the public library, without concern for whether the young person can or cannot find the information on his/her own.

## The Evaluability of Youth Services Library Programs

Rossi and Freeman define an evaluability assessment as, "A set of procedures for planning evaluations so that stakeholders' interests are taken into account in order to maximize the utility of the evaluation." (15) Using this definition, the audience(s) and potential uses for an impact evaluation of youth services programs must be specified--if for no other reason than to see if they are in conflict with one another.

## Users of Evaluation

Youth services librarians are overwhelmingly interested in evaluations which will prove that they deserve resource allocations proportionate to the importance of what they do, compared to other parts of the parent organization. Among other things, they want to figure out how to make themselves appear important to administrators.

Line administrators of youth services librarians want to understand, preferably in the unemotional quantitative terms they are familiar with, what it is youth librarians do, so they can either feel assured and be able to explain that resources are being used effectively, or so they can identify the need for possible resource reallocations, or so that they can use a favorable or unfavorable comparison of their youth library services with others for some political gain. Both levels of practitioners have very short-term, pragmatic, and local uses for an evaluation. They want data on which to justify their existence and their management decisions in a frequently adversarial administrative and/or funding climate. This climate does not foster time-consuming, bias-free evaluations.

The evaluation of federally funded library programs serves different uses entirely. Besides accountability--whether the money has been spent as proposed by the institution getting it--evaluators of federally funded projects want to know whether the project worked (internal validity); and whether it was worth replicating to see if it would work elsewhere (external validity); whether the program would or should continue in the absence of federal support. If the program did not work, they want to know whether continued federal support is likely to help make it work and within what time frame. If it did work, they want to know how the information could be diffused best to other practitioners and whether others would in turn, need federal money in addition to information to do it themselves. Besides these pragmatic uses, the U.S. Department of Education also needs to know whether national public policy objectives were met through the projects funded.

Given these differences in stakeholders' uses for evaluations, the historical and frequently inarticulate defensiveness of youth services practitioners, some means must be found at the outset of meshing the uses of evaluations of federal projects with uses at the local level. These needs mesh in five areas:

1) Money. The local library always needs it and the federal government needs to give it away. How to apply for a federal grant is a total mystery to most youth services practitioners at the direct service level. If there is no administrator at a departmental or district level to actually go after the money on behalf of the youth services staff, in most cases they will not or cannot apply. Since the funding priorities of the Library Services and Construction Act (LSCA) are not precise in terms of any age except older Americans, public libraries can easily ignore the fact that young people are a target audience, or that they constitute a dismayingly large share of the various disadvantaged audiences federal grants are intended to help. Moreover, libraries with money frequently are more adept at going after additional money than those without, whereas the latter libraries may need the money more and possibly do better programs with it. The entire application process needs to be simplified and publicized better.

2) Staff. It cannot be assumed by the federal grantor that additional work can be readily absorbed at the local level without some diminution in probable impact. If the local project director is to pay the necessary attention to implementing, supervising, and documenting the variables and processes of the program which relate to impact, the grant giver must assume some responsibility for providing both support staff and training grant managers. Librarians are notoriously undersupported in terms of clerical and administrative support staff, often in no place worse than youth services. If the program further involves preschools and/or nursery schools in any way, which is where most preschool

children are now, the grantor must acknowledge that their staffs are no better off than librarians in this regard.

3) Failure. The funded library must also be granted the right to fail to achieve the stated impact with impunity for good cause. The library, and more importantly, its governing body, must be assured that failure to achieve the project's objectives will neither embarrass them politically nor make them forever ineligible for future grants. Fear of failure can be further overcome by making it clear at the outset that a well-documented failure through which everyone learned something would be an advantage toward further funding.

4) Recognition. Federal grantors must take more responsibility for providing public recognition to the library programs they fund. Especially in public libraries, credit is needed for the local governing and funding bodies. Means of recognition can be as simple as a letter and framable certificate to hang in city hall, a special citation list in publications of the national groups of mayors and counties, or discretionary grants to local academics or free lance journalists to write up the programs in publishable form for government, library, and education periodicals. Federal funding priorities are often not politically palatable or in synch with those of local government. It is vital that the library program is not caught in the middle. A form of recognition which goes directly to elected officials is one way to get them to buy into the project and possibly protect the library. The one thing which is guaranteed to make elected officials happy is to credit them with helping children through their local library with outside money.

5) Technical assistance. Federal grantors must assume responsibility for training the local project staff in process evaluation techniques and for the impact evaluation itself. Impact evaluations are beyond the capability of most local youth services staff, not only because of interest bias and lack of technical skill in evaluation methods, but also because of the imprecise specification of short term expected impacts in the field of youth services. Librarians know what kind of young person they want to help produce, but the marks of progress toward that end are still operationally elusive.

Assuming this responsibility does not mean that the federal staff has to perform these activities, but they must pay for them. Good impact evaluations are labor-intensive, inherently political, and sophisticated. Just assessing the presence or absence of regression effects on a program is difficult. When the developmental vagaries of a target audience of children who cannot speak for themselves, and who want to please, or can be intimidated by adults are considered, in addition to multiple influences on multiple manifested impacts, an even greater need for evaluation expertise becomes apparent.

## Program Description and Documentation

Besides the stakeholder's potential uses of the evaluation, an evaluability assessment covers other facets of a program which must be present for an impact evaluation to take place. These include the rationale of the program, the documentation systems in place or needed to describe the activities of the program, determining the identity of the authority to commit resources and make modifications in the program, and the specification of the program delivery system. (15) It is helpful to discuss each of these facets in terms of youth library services, because a good argument can be made that most of these programs as established in public libraries are in a very primitive state of development in terms of evaluability, despite often intuitive brilliance in program design.

**Program Rationale.** While the long-term impact of youth services programs are relatively well defined, measurable intermediate impacts are not and there is little evaluation research from which to draw them. Fitzgibbons summed it well when she stated, "As in most areas of librarianship, research concerning services to children and young adults tends to be survey or historical, is usually non-cumulative, and tends to be unsophisticated in terms of statistical techniques." (16) Even the article previously cited by Loetscher, which documents professional activities in exemplary elementary school library media centers, is the result of a descriptive survey of what is there, rather than what might be. Because the schools have been called exemplary, the impact, as usual, is assumed. Even in terms of the most basic library process variables, DeProspo's performance measure research in the early seventies should have laid the idea of assumed outputs to rest. (17)

Ironically, one of the most ubiquitous programs in all public, and some school libraries, the summer reading club, has been researched peripherally for an intermediate impact which is measurable. Heyns, in studying the effects of schooling over the summer, found a relationship between reading scores and proximity to a public library. (18) A recent exploratory doctoral study to edify the process variables related to a successful summer reading club was completed by Locke. Forty-three percent of the respondents produced promotional materials with state assistance and 35 percent respondents selected the theme through the state library or through a statewide committee of librarians. (19) The maintenance of reading scores from June through September is logical, given the intended long-term impacts of elementary age programs, yet this impact is largely unexamined by evaluators, even by those in state library development agencies who contribute to the program theme materials and activities.

**Program Documentation.** Youth services programs are remarkably under-documented, especially in public libraries. In part, this is because the state of public library documentation is poor but improving, thanks to the effort of the Public Library Association (PLA), a division of the America Library Association (ALA), and also because of the sheer intransience of the youth services community over who can legitimately lay claim to youth as a target audience. Young adult librarians have always claimed them; children's librarians have always resisted, and nobody ever gets any further with the argument. The ALA Association for Library Services to Children (ALSC) defines a child as someone through 8th grade or age 13. The Young Adult Services Division (YASD) defines a young adult as "someone who no longer considers himself a child, but who society does not yet consider an adult." The most recent attempt to clarify the situation by a YASD Age Definition Task Force ended with an uneasy withdrawal by YASD, which continued the status quo. The first national survey of young adult resources and services in public libraries, conducted in 1987, used a working definition of ages 12-18 or 7th through 12th grades for young adults. This was verified by a validity question within the test of the survey instrument. (20) A new national children's services survey is underway, but the advisors to that study were not happy with the definition used in the previous one.

Local definitions are fine for local program evaluation, but they are useless for federal evaluators. Since it is quite obvious that the youth services community in public libraries cannot resolve this question on their own, state or federal intervention may be an unwelcome but necessary alternative.

Beyond the definition problem, one of librarianship's most cherished service ideals--the confidentiality of patron records--works against program documentation. Librarians are not used to, and generally uncomfortable with, asking program participants questions about themselves which are most helpful to impact evaluators. Except for gross demographic variables in amalgamated age categories, program participant information is neither routinely kept nor considered the library's business. To examine attrition as an impact variable is almost impossible in a public library setting, because there is no baseline data taken on program participants at the beginning to compare those who complete the program against those who drop out. I have experienced vociferous objections from colleagues on several occasions when designing reports or registration forms which asked for the age of the person in question. Evaluators of federal programs need to address the real and imagined fears of violating patron privacy to improve local documentation in public libraries. School libraries at least can document participation by class, grade, subject taught, etc., and attrition can be ascertained through attendance records, assuming accurate records and no prohibitions on an evaluator's access to them.

In those documentation systems which exist, besides the lack of an adequate operational definition of the target audience, youth services are ignored because of the sheer lack of uniformity in records kept. A recent survey of 80 children's services managers on the mailing list of the Children's Services Management Consortium in California, identified only four commonly kept measures which the Consortium requested state library reports mandate for documentation purposes. The absence of these four measures in the state reporting system for public libraries should be very revealing to federal evaluators. They are: 1) population of children age 0-14 years; 2) separate totals for library sponsored programs for adults and for children; 3) separate totals for attendance at library sponsored programs for adults and for children; and 4) separate figures for materials budget expenditures for adults and children. (21) The fact that these data are not kept and reported routinely is appalling, especially since an LSCA-funded feasibility study in Wisconsin on applying PLA's output measures methodology to children's services identified the need for them. (22) One small ray of hope for better documentation of gross input and output measures on public library children's services is the establishment of a joint ALSC/PLA Output Measures for Children's Services in Public Libraries Committee within ALA with a charge from both division boards to raise the money to adapt and enhance current PLA output measures and to create a manual for documenting them for children's librarians. (23) National impact evaluation could be greatly advanced by some federal contribution to this effort.

Impact evaluators must assume inadequate documentation systems at the outset of youth services program evaluation. An evaluability assessment will not only have to figure out what, if anything, is already in place, but will have to specify what needs to be created. This necessity implies that youth services program may need a longer start-up time than federal funders are used to, since little data will exist for the regular program, let alone that part to be specially funded.

**Program Authority and Delivery System**

Most youth services are delivered at a basic local level of service--a branch of a public library, or in an individual school building rather than a district office. Unfortunately, this is where the least empowered, and sometimes, the least qualified staff work. Since libraries and schools are generally very hierarchical, people are used to being told what to do. Since their needs for an evaluation are different from the needs of both their administrators and federal funders, it would be wise to assess the competence of the program delivery staff and involve them in both the program and the evaluation design at the outset, or run the risk of having evaluation work ignored or sabotaged. Federal funders would also be well advised to look at what support system the fund-seeking organization already has or proposes to put in place to help empower youth services personnel to do a good job.

The other thing which is very important to the successful implementation of a youth services program is the unambiguous support of top administration. Youth services people consistently feel undervalued, in large part because their service clientele does not have adult status and are actively loathed or only tolerated by colleagues. In schools, librarians are frequently regarded as instructional interlopers who have opted out of the classroom for an easier life. It often takes this unambiguous public support by administrators to suppress the annoyance of colleagues whose support, even grudging at times, is needed. Just listening to circulation staff complaints in public libraries during the first weeks of a summer reading program can be enlightening in this regard. This syndrome just adds another layer of paranoia to the overworked staff serving youth and the program evaluation should be designed, insofar as possible, to defuse it.

**Beyond Evaluability.** Mandating an operational definition of child and young adult and requiring routine statistical reports on service for this clientele at the state or federal level will advance almost any kind of evaluation of youth services in libraries because the accumulation of a baseline service database will finally begin. There are, however, some other activities which might advance the impact evaluation of youth services programs:

1) Create a cadre of local and regional evaluators.

The evaluators for local youth programs do not have to be either the people actually doing the program or state library agency personnel. Most states have either a library school or a university with an appropriate academic department with faculty knowledgeable in research methods who want tenure. There are few places where library school faculty teaching youth services can either get published in scholarly journals or get outside money to fund research on youth services. The U.S. Department of Education should consider identifying and subcontracting with these people through their departments to do impact evaluations and evaluability assessments at the local level. The Department could even adopt the National Endowment for the Humanities (NEH) model of teams of paired practitioners and academics as part of the proposal process, managing evaluator input to program design and documentation systems at the time the program is being proposed for federal funding.

The academic evaluators would: a) Get access to program data for analysis without having to write an original research proposal themselves; b) Have to write an evaluation report for the U.S. Department of Education which would force them to organize the material in a form easily turned into a subsequent scholarly publication; c)

Bring extra funding into their departmental budgets to make them more valuable as academics who can get outside money; and d) Get richer, research based, practical examples for classroom teaching.

2) Dedicate LSCA dollars to youth services programs in public libraries.

In public libraries, more impacts of federal programs could be assessed if more programs existed. While a separate exclusive title dedicated to funding for youth services within LSCA is ideal, it is probably unlikely, given the political climate within which LSCA reauthorization is considered. In the absence of some sort of exclusive prioritization, the potential of requesting LSCA funding for youth services programs should be made clear in the guidelines for proposals. Youth services practitioners are convinced by experience that children and young adults as a possible target audience must be made explicit, or their implicitness will be ignored. Saying that other generalized LSCA titles include children and youth simply does not work, especially when the public library community most adept at grant proposal writing are not youth services practitioners. Future guidelines for LSCA proposals must specifically name children and young adults as possible clientele.

3) Create knowledgeable youth services advocates in state library development agencies.

The recent survey of young adult services, sponsored by the U.S. Department of Education's National Center for Education Statistics, revealed that the majority of public libraries serve adolescents through a generalist staff with no access to specialty training, even through 25 percent of their clientele is in this age category. If developmentally sound, fundable programs with a good chance of impact are to be designed, it is unlikely to come from these librarians without some other intermediate intervention. If the U.S. Department of Education wants such programs, some attention to the erosion of youth services positions in state library agencies must take place. Neither the Department nor the library community at large can continue to depend on a dwindling supply of youth services specialists to bail them out of a general miasma of their own making. Specialists are more expensive, but if their existence is tied conceptually and practically to the production, identification, and assessment of impacts on the lives of the next generation who, not so coincidentally, will be supporting the author and readers of this paper in their old age, they would seem to be well worth the money.

**Action Agenda**

Youth services library programs in general are not in good shape for easy impact evaluations. While the long term impacts are generally agreed upon, scoring the library program's unique contribution is still problematic. Besides the almost endemic suspicion youth services practioners have toward outside evaluators, there are no agreed-upon operational client definitions, poor service documentation at local and state levels, little technical skill in evaluation methodology, and imprecise specification of intermediate impacts. Federal program evaluators need better people in place locally for both staff development in youth services and for the evaluation of youth services programs during and after the period when they are developed. The uses of evaluations must be considered in advance so that the interest of all stakeholders in the evaluation are acknowledged and represented. In youth services especially, the uses are quite different from those of administrators or federal funders. The U.S. Department of Education must address means of improving the evaluability of youth programs in general before impact evaluations will be possible.

## Notes and References

1. Carlson, Ann D. *Early Childhood Literature Sharing Programs in Libraries* (Hamden, CT: Shoe String Press, 1985).

2. Kontos, Susan. "What Preschool Children Know About Reading and How They Learn It," *Young Children* 41 (November, 1986): 58-66.

3. Schickedanz, Judith A. *More than the ABCs: The Early Stages of Reading and Writing* (Washington, D.C.: National Association for the Education of Young Children, 1986).

4. McClure, Charles R. and others. *Planning and Role Setting for Public Libraries* (Chicago: American Library Association, 1987).

5. Loertscher, David V. and others. "Exemplary Elementary Schools and Their Media Centers; A Research Report." In: Frances Beck McDonald, Compiler. *The Emerging School Library Media Program: Readings* (Englewood, CO: Libraries Unlimited, 1988).

6. Rollack, Barbara T. *Public Library Services for Children* (Hamden, CT: Shoe String Press, 1988).

7. Hopkins, Diane McAfee. "Elementary School Library Media Program and the Promotion of Positive Self-Concepts: A Report of an Exploratory Study," *Library Quarterly* 59 (April, 1989): 131-147.

8. Edwards, Margaret A. *The Fair Garden and the Swarm of Beasts: The Library and the Young Adult.* (New York: Hawthorn, 1974).

9. Greenberg, Marilyn W. "Developmental Needs of Young Adults and Library Services," In: JoAnn V. Rodgers, Editor. *Libraries and Young Adults: Media, Services and Librarianship* (Littleton, CO: Libraries Unlimited, 1979).

10. Chelton, Mary K., "Developmentally Based Performance Measures for Young Adult Services," *Top of the News* 41 (Fall, 1984): 39-52.

11. Liesner, James W., "Learning at Risk: School Library Media Programs in an Information World," *School Library Media Quarterly* 13 (Fall, 1985): 11-20.

12. Hirsch, E.D., Jr. *Cultural Literacy: What Every American Needs to Know* (Boston: Houghton Mifflin, 1987).

13. Printz, Mike, Librarian, Topeka West High School, Topeka, KS. Telephone conversation, April, 1989.

14. Spencer, Pam, Librarian, Thomas Jefferson High School of Science and Technology, Fairfax County, VA. Telephone conversation and correspondence, April 1989.

15. Rossi, Peter H. and Howard E. Freeman. *Evaluation, A Systematic Approach*, 3rd ed. (Beverly Hills, CA: Sage, 1985).

16. Fitzgibbons, Shirley, "Research on Library Services for Children and Young Adults: Implications for Practice." In: Ken and Carol-Ann Haycock, Editors. *Kids and Libraries, Selections from Emergency Librarian* (Vancouver, B.C.: Dyad Service, 1984).

17. DeProspo, Ernest R. and others. *Performance Measures for Public Libraries* (Chicago: American Library Association, 1973).

18. Heyns, Barbara. *Summer Learning and the Effect of Schooling* (New York: Academic Press, 1978).

19. Locke, Jill L. "The Effectiveness of Summer Reading Programs in Public Libraries in the United States." Unpublished Doctoral Dissertation, University of Pittsburgh, 1988.

20. U.S. Department of Education. Office of Educational Research and Improvement. National Center for Education Statistics. *Survey Report: Services and Resources for Young Adults in Public Libraries*. (Data Series: FRSS-28) (Washington, D.C.: U.S. Government Printing Office, 1988).

21. Parikh, Neel. Letter and attachments to Gary Strong, September 8, 1987.

22. Zweizig, Douglas L. *Output Measures for Children's Services in Wisconsin Public Libraries: A Pilot Project--1984-85*. (Madison, WI: Division for Library Services, 1985).

23. Reif, Kathleen, and Clara Bohrer. Memo to Eleanor Jo Rodger and Susan Roman, October, 1988.

# QUALITATIVE AND QUANTITATIVE EVALUATION: EIGHT MODELS FOR ASSESSMENT

**Betty J. Turock**

## Abstract

Librarianship has insisted on promulgating one method of evaluation at a time--first, the use of input measures, then the use of output measures--even in the face of evidence that they may make difficult the fair assessment of nontraditional library programs in non-traditional settings. This paper identifies and discusses eight quantitative and qualitative models for the valid, reliable assessment of federally funded public library programs, pointing out their strengths and weaknesses. Advising that we should choose from among them depending on situational contexts and contingencies, the author exhorts us to diversify.

It is no secret that we are operating in what MIT economist, Lester Thurow, has called a Zero Sum Society. (1) Our country's slowed growth in productivity in a world market filled with competitive peers and our mounting budget deficit has resulted in a decline from our previous global economic superiority. Both Japan and West Germany have surpassed us. Although America's $2 billion surplus once set us up as the leader in world markets, we now have a trade deficit. While in the past we had net foreign assets of $152 billion, in 1989 we are among the world's largest debtor nations.

There can be little doubt that these economic facts will ultimately effect the financial well-being of United States public programs, including its libraries. Accountability and productivity have become national bywords and if, as in the past, trends set in motion by the federal government ultimately pervade state and local practice, an even greater emphasis on this duo lies ahead.

We must make more visible the magnitude of the public library's contribution to our society, if we are to successfully compete in funding arenas. The U.S. Department of Education's Office of Library Programs will need hard data to present at Hearings for Congressional Committees struggling to correct the economic ills of the country--data state library agencies can help supply through improved evaluation of federally funded library programs.

## Evaluation Neglect

We are a profession that has been exhorted to improve evaluation for over two decades. Still, when the issue is clearly not whether, but how to evaluate, we, not unlike other professions, have resisted.

Perhaps part of out neglect arises from confusing monitoring with evaluation. Monitoring, which is undertaken by federal program officers, involves tracking for control and compliance for regulatory purposes. Now, frequently, monitoring activity is limited to oversight of the fiscal aspects of federally funded programs and that is all the assessment that occurs.

Evaluation, on the other hand, entails formulating questions about the efficiency and effectiveness of the programs in question; adopting and defining standards by which to judge success; selecting designs and sampling procedures to supply a read out on programmatic activity, including identifying the measures set to gather data; collecting information on specified program elements; analyzing the information gathered to determine the degree to which the program meets the standards; and, finally, reporting results to evaluation stakeholders. (2)

Perhaps part of our resistance also comes from our perception of evaluation. At its best, evaluation is a judicious critique, meant to improve, not, as is commonly held, to prove or disprove. The means and methods of evaluation are set forth in planning and proceed in a cycle of planning-evaluation-planning that is continuous, allowing us to alter the course of assessment if it proves insufficient to the task at hand. Summative as well as formative evaluation should have improvement as its goal, especially when program adoption by different libraries follows.

### Program Evaluation: More Than the Application of Output Measures

Our profession has insisted on promulgating one method of evaluation at a time--first, the use of input measures, then for more than 15 years, the use of output or performance measures, introduced to public libraries through a grant from the U.S. Office of Education, Bureau of Libraries and Educational Technology to Ernest R. DeProspo. (3) That work was expanded in *Output Measures for Public Libraries* (4), most recently revised in 1987. (5)

We have found evidence that output measures mitigate against documenting high performance in libraries situated in economically disadvantaged communities. Fairness dictates that we must find a less biased, more objective means of assessing these libraries and their programs. Currently, we seldom even encounter reference to measures, such as equalized valuation per capita, that can provide the basis for comparing actual fiscal support with the associated community's economic ability to supply that support.

### Evaluation Essentials

As Ernest R. House puts it, "At its simplest, evaluation leads to the settled opinion that something is the case. It does not lead to a decision to act in a certain way, that entails administrative judgement." (6) Instead of being cut from funding, programs that make a real effort to improve, could be continued.

To be meaningful, just, and true, an evaluation must be developed through dialogue. If it is to be accepted by those evaluated as well as by those who might later assess the evaluation, the process, and the product must meet all three standards. The evaluator and those evaluated have to agree that criteria applied to decide success can fairly determine what is going on in the environment, as well as what is going on in the program. Definitions for measurement must be agreed upon in advance and employed in the same manner if findings are to be subjected to comparison and the criteria weighed and balanced as those who know the situation best would weigh and balance them. The aim of evaluation is to express the truth, not only for the library and the librarian, but also for the citizens the program was designed to serve.

Just as there is a difference between the rigorous nature of the evaluation needed for formative and summative purposes, there is also a difference between the evaluation conducted on programs funded at high and low levels. Half of the funds awarded should not be spent on evaluation to provide the valid scientific proof, an illusionary goal at best.

Evaluation, regardless of the size of the program or its scope, is situational and dependent on the contingencies of the site from which it is operating. As Michael Patton has noted, "Evaluation emerges from the special characteristics and conditions of a particular situation--the mix of people, history, context, resources, constraints, values, needs, interest, yes and chance." (7)

So evaluation solicits more than numbers; it solicits information about organizational dynamics, and environmental uncertainties to answer questions of impact. At the same time, public library program evaluation is constrained by political realities, the frustration of too little money and time, and the demands from funders for assessments that play a meaningful part in programmatic development. Still evaluation cannot be avoided or ignored; it is always part of program planning where continuous improvement and library development are the ultimate goals of funding.

## Mixing and Matching the Ubiquitous Eight

Ernest R. House has identified eight models which could supply the framework for the valid and reliable assessments which have eluded librarianship. (8) Their comparative characteristics are shown in the chart below.

### EVALUATION MODELS AND THEIR CHARACTERISTICS

#### SYSTEMS ANALYSIS APPROACH

Major Reference Group - Managers, particularly of federal programs
Outcome - Efficiency
Methodology - Cost/Effectiveness, Cost/Benefit Analysis
Typical Question - Can the results be produced more economically?

#### DECISIONMAKING APPROACH

Major Reference Group - Decisionmakers, especially policymakers
Outcomes - Utilization of Evaluation Results
Methodology - Surveys: Questionnaires, Interviews
Typical Questions - What decisions will this evaluation help make?

#### GOAL-BASED APPROACH

Major Reference Group - Managers
Outcomes - Accountability
Methodology - Program Objectives
Typical Questions - Is this program achieving what it intended?

#### GOAL-FREE APPROACH

Major Reference Group - Program Consumers
Outcome - Consumer Choice, Social Utility
Methodology - Bias Control, Logical Analysis
Typical Questions - What are all of the effects on the client?

### ART CRITICISM APPROACH

Major Reference Group - Connoisseurs, Consumers
Outcomes - Improved Program Standards, Heightened Awareness
Methodology - Critical Review
Typical Questions - Would a critic approve this program?

### PROFESSIONAL REVIEW APPROACH

Major Reference Groups - Professionals, Public
Outcome - Professional Acceptance or Agreement
Methodology - Panel Review, Self-study
Typical Questions - How would professionals rate this program?

### QUASI-LEGAL APPROACH

Major Reference Groups - Jury
Outcomes - Resolution of Controversy
Methodology - Quasi-Legal Procedures
Typical Questions - What are the arguments for and against the program?

### TRANSACTION APPROACH

Major Reference Groups - Community Members
Outcome - Understanding Diversity
Methodology - Case Studies
Typical Questions - How does the program look to different people?

These quantitative and qualitative models are brought together here so that librarians can throw away any notions perpetrated by misguided methodologists that numbers and words cannot coexist in the same evaluative process, or that systematic qualitative approaches do not result in evaluations equally as valid and reliable as those from quantitative sources.

The models are presented in order of the prescriptive nature of their methodology. As we proceed through the list we see movement from the quantitative to the qualitative. Depending on the level and specificity of the evaluation contemplated, the models can be used singly or in combination. So that evaluative questions can be weighed vis-a-vis the complexity of the means employed to discover some answers, the strengths and weaknesses of each model are pointed out.

### The Systems Analysis Approach

Advocates make the case that this is the only scientific route for evaluation. They claim that by following rather precise quantitative guidelines, reliable, hard data are produced. To date the predominant problem which this approach has addressed is measuring the outcome of government programs.

Program evaluators who use the Systems Analysis model collect data on a few defined output measures deemed critical, such as in-library circulation, out-of-library circulation, and others, determine costs resulting from their production, and associate differences in program or policy outcomes to variations in the measures. Generally higher scores are interpreted as meaning greater success. The relationship of outcome measures to program achievement is demonstrated through statistical techniques like correlational analyses. The programs determined most efficient have the highest possible activity measurement at the lowest possible cost.

Since the purpose is to establish cause and effect, good experimental design is preferred. The randomized control group, although the design of choice, is not always possible, and quasi-experimental methods are frequently invoked. After outcomes are measured, programs are compared on costs to determine which outcome is produced at least cost; so a major goal of the approach is to tie input to outcome.

Many of these evaluations use test scores as the only measure of success. They are compared to normative data gathered on large numbers of cases over a long period of time. A comprehensive evaluation based upon the Systems Analysis Approach answers questions about:

> Program planning by supplying information that helps future planners identify appropriate programs for specific social problems;
>
> Program monitoring by highlighting whether a program is in conformity with its initial design;
>
> Impact assessment by demonstrating whether the program has produced changes in the desired direction;
>
> Economic efficiency by providing information that indicates whether the program is efficient. (9)

While this approach leads to data that gives the government specifies on outcome questions asked in assessments, it can exclude the interests and concerns of program participants, particularly those in the lowest rungs of the social hierarchy. Objectivity is the first priority, but in reducing everything to a few indicators to demonstrate reliability, do cost-benefit analyses, and discover the most efficient programs, the outcomes of complex social programs may be narrowed to what is quantitatively measurable.

When the range of data collected and the context are given minimal attention, these assessments are often not credible to those evaluated. One way of strengthening the Systems Analysis Approach is to broaden the number and types of indicators used.

**Library Applications.** In librarianship, one of the early efforts to follow the Systems Analysis Approach for evaluation was produced in 1974 by Morris Hamburg and his associates at the University of Pennsylvania who developed a single overall measure of public library performance. (10) When they concluded that the major function of libraries was to expose people to records of human knowledge, they proposed item-use hours as the basic measure of library outcome. Library use for whatever purpose--circulation of materials, satisifying reference questions, etc.--was translated into user time in contact with documents, then summed across services for one total.

In the main, attempts to identify single measures of effectiveness have given way to attempts to identify the multiple indicators and dimensions of effectiveness. (11) But as recently as 1982, Daniel O'Connor proposed one score, the Library Quotient (LQ). (12) His method is based on standard scores to evaluate public libraries, patterned after similar scores widely used in educational achievement and intelligence testing. This process could be applied to the evaluation of federally funded programs as well.

O'Connor converted data into ratios for proportion of the budget spent on materials, new volumes per capita, patron visits per capita, reference visits per capita, and in library use of materials per capita and transformed them into standard scores where each library's position on any performance measure was a function of the positions of all other libraries.

The standard, or z-score, became the basis for an LQ, which was claimed capable of providing a national comparison of all public libraries for evaluation and planning purposes. O'Connor's point was not that input scores should supplant output scores, but that input scores can serve important purposes, like productivity measurement, by relating benefits to costs. He suggested that excellence in performance could be identified by specific cutoff points in the continuum of LQ scores.

The development of national standards for public libraries using standard scores would of necessity be based either on a large randomly selected national sample of libraries or on the entire universe of United States public libraries.

While there are no normative databases at present, the newly initiated Federal-State Cooperative System for Public Library Data (FSCS) can be built to provide such information in the future, since it is being gathered over the universe of public libraries in the United States. (13) The database might be structured to include a corporate memory of evaluation results which were related to other library data as a means of refining and improving the evaluation process.

For the evaluation of federally funded library programs the Systems Analysis Approach works best at the summative level, particularly when programs are attempting to prove they are exemplary, or for program evaluations that can compare participants' pre- and post- program scores to standardized scores, empirically demonstrating the extent of the program's effects.

Literacy is the LSCA priority that comes to mind immediately where this technique is appropriately adapted. Since there are numerous valid, reliable, standardized tests of reading achievement, before and after scores for literacy program participants would provide strong evidence of program effectiveness and, when compared to library costs, in most cases program efficiency.

**Decision-making Approach**

Since, for those who support the Decision-making Approach, evaluation is aimed at action and change, the evaluative process must have as its priority production of information that will affect policymaking while it improves effectiveness. Research has documented that program initiators and implementors frequently have little faith that evaluation will actually impact decisions about programmatic futures. (14) The common belief grows that even if performance data were collected conscienciously, going to the bargaining table with that data would not lead to rewards commensurate with the work involved. The importance of affecting contemporary government decision makers through evaluation is underscored by the fact that, while chairing the Joint Committee on Standards for Educational Evaluation, Daniel L. Stufflebeam led the fight to list utility as the first of the primary characteristics for exemplary evaluation.

With utility as the hallmark, then, the Decision-making Approach begins by identifying intended evaluation users and stakeholders. The consequence of lack of participation is assessed for each group of stakeholders not involved. If their inclusion is crucial, every effort is made to gather them into the process.

Following a stakeholder survey, the evaluation focuses on their questions, issues they raised, and how they intend to use any answers generated through assessment. The decision-making situation is projected and the evaluative criteria for each decision are spelled out. The appropriate methods for data collection and analysis flow from the decision-making issues uncovered. When the component parts of a program are identified and the possible means of evaluating them attached to the decisions on which they will bear relevance, decision makers rate the information they will receive from each evaluative approach. The utility of the action alternatives are summarized across the decision makers' value dimensions to determine which course has the greatest utility for the evaluation. Until an unequivocal "yes" is returned to the question, "In view of the use expected, is the evaluation worth doing?" design remains incomplete and successive iterations of fact finding are run.

Since the decision makers are the audience for the evaluation, their concerns and criteria are considered most significant. The decisions made at the top of the hierarchy--usually revolving about the effectiveness of the program on some dimension, and in particular, which parts of the program are working--receive the thrust of the assessment. Quantitative data are the common source used to demonstrate effectiveness. But while utility of information is an important criterion for evaluation, it is not the only one; a useful criterion must also be true and fair.

A great advantage of the Decision-making Approach, on the other hand, is that it underscores the practicality of evaluation. The information that is collected is the information that is most useful to decision makers. Obviously, credibility of the data is high for the intended audience. To overcome the use of evaluation in an unethical way, the decision-making groups must be defined broadly. The great danger for the evaluator to guard against is becoming the decision makers' pawn.

**Library Applications.** At the close of a federally funded program for older adults the Board of Trustees may have to decide whether or not to continue the service initiated by a grant under LSCA Title I. At the same time the President of the Board of Trustees wants a political career and one of the criterion affecting his decisions about library programs is whether or not they will increase his visibility in a positive way among his possible future constituents. Interviews with the Board members and other stakeholders will form the basis for formulating questionnaires that design an evaluative study speaking to the articulated information needs for decision making. It is important to get below the surface and determine the real information sought. For example, in evaluating the program serving older adults, information about the number of voters among elder participants might be as important as information about the number of elders who take part in the program.

## Goal-Based Approach

The most familiar and the most popular among evaluators, this model is also currently the most commonly put forth idea for evaluation. Like the Systems Approach, it relies on quantification and follows "scientific" procedures.

Here the identifying feature is the presence of some goal or objective whose measure of attainment constitutes the main focus of the evaluation problem. This approach takes the goals and objectives as stated and collects evidence to determine whether the program has achieved them. The goals and objectives are the criteria against which the evaluator assesses what the program developers said they intended to accomplish; in effect, they serve as the exclusive source of program standards. The purpose of the evaluation is to measure the level of achievement of desired effects against the goals that were set out as a means of contributing to subsequent decisions about the program.

The evaluation process is conceived of as setting goals and objectives; identifying goal activity; putting goal activity into operation; assessing the effect of the goal operation, including value formation; and goal measuring. A clear program objective is equivalent to the hypothesis in a research study. The primary methodology centers on collecting field data on quantified variables and on the means of measuring success.

The difference between the Systems Analysis and the Goal-Based Approaches is that, in the first, only a predetermined number of criteria are tested and it is assumed that these criteria are critical. In the second, evaluation would determine whether each objective is achieved. So success is described in terms of prespecified and measurable objectives rather than in terms of values actually obtained.

Validity is derived from holding the program accountable for its prespecified goals, but the approach does not include methods for judging the correctness of the goals themselves, nor does it ask whose interests the program represents or whether important outcomes are neglected by the prespecification of objectives. Proponents stress the accountability aspects of the model, since program claims are frequently the basis upon which public funds are awarded.

**Library Applications.** Not unexpectedly, the Goal-Based Approach has supplied most of the framework for the contemporary evaluation of library programs and library performance. The Public Library Development Project from the Public Library Association, a division of the American Library Association, and its resulting products, *The Planning Process for Public Libraries* (15), Planning and Role Setting for Public Libraries (16), along with *Output Measures for Public Libraries* first and second editions, proceed from this frame of reference. The extension course, entitled, "Are We There Yet?" developed by Jane Robbins-Carter and Douglas Zweizig and offered by the University of Wisconsin School of Library and Information Studies through American Libraries (17) provides a step-by-step outline for implementing a Goal-Based evaluation.

This model is a natural candidate for the evaluation of LSCA Title I programs. For example, a program might have as its goal improving services to the physically handicapped. An objective might be to reach 10 percent of the physically handicapped population in the library's service area in the first six months of operating a new Media Home Delivery Service. The evaluation would measure what has occurred against what was established as the target.

## Goal-Free Approach

Created in direct reaction to the ubiquity of the goal-determined evaluation, the Goal-Free Approach remains one of the most talked about and least used processes. Developed to reduce bias, it requires an outside evaluator to carry it out.

The evaluation is not based on program goals. In fact, the evaluator remains uninformed about them and searches for all program outcomes, many of which are side-effects or unintended results, both positive and negative. In this case, the evaluation is not looking for intention, but achievement.

Among the models previously presented, the traditional notion of objectivity is built on quantitative assessment alone, but the goal-free notion of objectivity is qualitative. Consumer Union uses this approach by focusing on criteria they think will benefit consumers rather than using the producers' goals. Bias is reduced by lack of knowledge of the overt program goals and independence from program personnel. Side effects, which are downgraded in other approaches, have equal weight here. Needs assessments become the source of standards.

While this model reduces the bias of searching only for program developers' intents, the evaluation's credibility is often called into question because external audiences along with those evaluated do not believe that the evaluator fully understands and appreciates what the program is doing.

Unlike the Goal-Based Approach, the Goal-Free Approach lacks a highly specific technology for evaluation. The evaluator functions like a detective armed with a variety of techniques but not with lock-step procedures. Observation is one of the major techniques employed. Running notes are made at the time of the observation, field logs are constructed later. Chronologs run events along a time line; context maps provide locale sketches and diagrams; sociometrics supply the information on interactions and relationships. Rating scales are constructed, derived from audience needs, which are used to rate crucial elements of the program, make an overall assessment of what it has accomplished, and the value of that accomplishment for the intended audience.

**Library Applications.** The Goal-Free Approach would be a good model for evaluation in the case of many LSCA Title I priorities. For example, it could be applied to determine the success of a program funded to provide materials supporting afterschool reading. If reading were meant to increase skills by exposure to a wide range of high interest, low ability materials, a number of indicators might point to the merits of the program. An examination of pretest scores of students; visiting scheduled tutoring sessions; interviewing tutors and students; reading expert reviews; and examining the materials themselves would provide abundant data that could substantiate success or failure. By developing a rating scale pointing out the characteristics of the materials required for a program of this type, a panel of judges--students, teachers, librarians, and parents--might decide the adequacy of the materials purchased on the scale; a combination of their scores might provide an overall assessment of the materials element of the program.

## Art Criticism Approach

This is another newer qualitative approach. Evaluators draw on their own experiences and intuitive reasoning to judge what is happening and to express their judgments in language and concepts that nonexperts can understand. Some questions that the Art Criticism Approach seeks to answer include: Are the people for whom the program was designed being helped? Are they acquiring habits conducive to their further development?

Like art critics, evaluators find themselves with the task of rendering the essential qualities constituting works of art, or excellent programs, into a language that will help others perceive the work with greater awareness. Judgments are based on the evaluator's own derived standards of excellence. Criticism is always qualitative. It is not negative appraisal, but rather the illumination of characteristics so that value is perceived. The critic has the experience to be able to distinguish what is significant. Proper training and experience are necessary for this connoisseur to make evaluative discriminations.

The consequence of the Art Criticism Approach is the development of connoisseurship in others. This is especially important in new areas of service where experts are just beginning to develop. The evaluator will find out whether the program is a good one, but the evaluative report will also heighten the awareness about what constitutes a good program. The critic-evaluator renders a situation to significant aspects of the program and captures its essence by presenting feelings as well as facts about its merit.

One of the problems in using the Art Criticism Approach is the tendency to pick critics who think as we do. Evaluators who use the Art Criticism Approach should make clear the values they hold so that the evaluation can be judged for bias and fairness within that context.

While there is no standard methodology for the Art Criticism Model except critical review, it is implemented in a couple of fairly standard ways. Immersion and familiarity with the program are vital. Referential Adequacy is frequently used to establish the validity and reliability of the evaluative critiques, that is, as observations are retained in notes, video tapes and similar recorded materials, a portion of the data is archived and not included in the initial analyses. Later it serves as a benchmark against a second data analysis and interpretation. The archived materials are recalled when tentative findings have been reached and referential adequacy is sought, that is, the data are analyzed to determine if features to which the critic pointed in the initial analysis can be found in the data retrieved from the archive. The second data review is used to demonstrate that different analyses reach similar conclusions. Skeptics can cull through the materials to satisfy themselves that the findings and interpretations are meaningful by testing them, directly.

**Library Applications.** This approach would provide a good option for application to the evaluation of newly burgeoning library programs for latchkey children. Evaluators would have been immersed in reading, writing, teaching, and other evidences of the problems and the lives of latchkey children. They would be familiar with library programs considered exemplary across the country and the elements that lead to success. The evaluators could be described as connoisseurs of the subject. The review of the specific program and the expressed judgments would inform and educate those evaluated and/or less knowledgeable.

### Professional Review Approach

Here evaluation is conducted by a team of peers who are assumed to have qualifications to judge the professional merit of a program. Procedures vary, but the evaluation culminates in a holistic assessment of the program by other professionals. Paul Dressel's overview of the self-study process details its organization, execution, and results. (18)

Before evaluators visit the site, the program staff engages in self-evaluation. They are appointed to self-study committees that review each of the program's functions and prepare a program profile. A comprehensive self-study is composed of data collection, assessment of strengths and weaknesses, re-examination of goals, and a detailed analysis of present and needed program resources. The self-study turned over to the peer reviewing team includes: definition and clarification of program purposes and goals; examination of the adequacy of physical and financial resources; study of the effectiveness of governance and decision-making processes, including roles of various groups; appraisal of the quality, morale, and activities of program staff; review of the strengths and weaknesses of the current program organization and delivery methods; consideration of the overall program climate and environment, including the role of the users, their satisfactions and dissatisfactions with the program and its services; and, finally, a collection of evidence on the effectiveness of the program and the process of user development.

The staff selects the peer review panel to validate the self-study. The program profile data are distributed and explained to the review panel by those responsible for their collection. The peer review panel members indicate the differences in their evaluation from the staff review before they leave the site, give a brief oral report pointing out the strengths and weaknesses of the program, and make recommendations for change. Program certification is awarded or withheld on the basis of the panel's report. After the visit, the program is expected to correct perceived weaknesses.

Program certification indicates to the general public, to the government, and to other institutions and similar programs the presence of at least minimal qualifications for accreditation. The recent trend for the public to challenge the right of professionals to control their own affairs underscores the belief that professions will not police their own operations very vigorously. This is counterbalanced by the result of public pressure for accountability which has increased the number of professional reviews undertaken each year.

**Library Applications.** Currently Leigh Estabrook is leading renewed discussion about accrediting public libraries using the Professional Review Approach. (19) If a general assessment of public libraries were to occur, an accreditation process would be played out. First, agreement would have to be reached on what constitutes an excellent library. Those criteria could be organized both for the entire library and for its functions--reference, adult services, children's, etc. A checklist of criteria would be prepared as a guide for evaluating each of the functions.

The model could also be applied to LSCA evaluations, especially in formative stages of program development. In evaluating an adult basic education program operating out of a library, funded under LSCA Title I, one of the criterion for excellence might be that, "attention is given to improving study skills." The review panel might place that item on a checklist and mark the quality or extent of what they found on a five-point scale from missing to excellent. Each of the major program functions would have similar checklists where criteria would be evaluated. The checklists would be totalled for a holistic evaluation of the program.

### Quasi-Legal Approach

Blue Ribbon Panels, like the Kerner and other Commissions, fall within the Quasi-Legal Approach. Presidentially appointed, members of the Panel heard evidence from witnesses, conducted their own investigation, and came to conclusions about civil disobedience in the United States. Based on the supposition that the facts in a case can be uncovered best if each side strives as hard as it can, in partisan fashion, to bring the most favorable evidence for its view to the attention of the panel, this model usually addresses controversial issues to resolve doubts about them.

This is a qualitative approach which employs a method typical of public hearings and mock trials. The aim often is to resolve the issue of program merit one way or the other. Evidence is presented by advocates to prove that the program is worthwhile and by adversaries to prove that it is ineffective. Most frequently, two teams battle over the summative question of whether a program should be continued or over a decision about renewed funding.

The approach is patterned after the courtroom; it reveals vital evidence rendered before a tribunal. Witnesses testify to submit the evidence. Rules are formulated about who may testify and the conditions for the testimony. Evidence includes not only facts, but also perceptions, opinions, biases and speculations.

The Quasi-Legal Approach has four stages: Issue generation, where sometimes as many as 30 or more interviews are conducted; issue selection, where surveys are circulated to reduce the issues to those that are crucial; argument preparation; and the hearing. (20)

The strength of the approach is that it incorporates the procedures and authority of law. But there is no body of case law by which to decide issues on the basis of precedence; each case is unique. Its major advantage is that pressing public issues can be addressed quickly by the appointment of a panel who bring the issues to an early resolution. Participation in the process is usually very broad and includes groups that might be excluded by most other approaches. A major appeal is its potential openess to diverse viewpoints.

**Library Applications.** Clearly, the approach has promise for LSCA Title I programs, some of which are innovative and may tend to generate controversy. For example, in a decision about whether or not to continue to fund a library-based career center, members of the Blue Ribbon Panel, appointed perhaps by the State Library, could interview key members of the staff to ferret out critics and their issues. They would develop a questionnaire and send it to a broad number of library stakeholders, including administrators, persons served, and government officials in addition to staff members, to gather opinion. Arguments would be prepared for and against the continuation from the data and the opinions of the partisans. A hearing would take place before the panel and a decision would be made by the members immediately following its conclusion.

### Transaction Approach

This model, which uses the qualitative Case Study for data collection and analysis, is becoming increasingly popular. It provides a way of judging programs within the context of their environment. Rather than pushing for quantification, this model pushes for understanding. Its strength lies in its ability to assist us in determining how to create programs that are responsive to nontraditional audiences.

The aim of the Transaction Approach is to show how a program is perceived by diverse groups. Here the evaluator arranges for different persons, representing sometimes disparate groups, to observe the program and assist in its evaluation.

The Transaction Approach is similar to the Art Criticism Approach, but it differs from the latter in that critics rely on their own experiences and apply standards of their own choosing. In the Transaction Model, evaluators report on the perception of others as well as their own in giving their judgment of a program. In the main, the Art Criticism Approach is best applied to formative evaluation and the formalized Case Study methodology is best applied to summative evaluation. Since it attempts to improve the understanding of the audience, the program staff, and sponsoring agencies about the program and what is going on in it, this model collects data to demonstrate how the program is perceived by others, particularly by the audience it was intended to serve.

In the past the Case Study has had difficulty establishing its credibility in a predominantly scientific community, but over the last 10 years, rigorous but flexible qualitative procedures for its execution, developed and described by Yvonne S. Lincoln and Egon G. Guba (21), Mathew B. Miles and A. Michael Huberman (22), and others have led to greater acceptance.

The Case Study concentrates on the description of program processes as well as outcomes. Program observers prepare and submit narratives, portrayals, and graphics for member checks. Evaluators find out what is of value to program audiences and gather expressions of worth from various individuals whose points of view differ. They check the quality of the records, get program personnel to react to the accuracy of their portrayals and audience members to react to the relevance of their findings. Methodological consistency and interpretation remain the primary problems in using the Case Study Method and the Transaction Approach. On other hand, they provide rich and persuasive information that is not available from other models.

**Library Applications.** There is no method that gives better results for the evaluation of new or innovative programs meant to reach nontraditional library audiences than the Transaction Approach. One example where a Case Study application would be meritorious is in the evaluation of a program for high school drop-outs that has as its goal providing nontraditional means to earn a high school diploma. Since, other than circulating self-study GED manuals, the library has had little experience in this area, the full range of qualitative methods could be applied to bring a better understanding of what is needed to make such programs successful and provide for their transportability to other library locations. Qualitative methodologies are systematic and rigorous, not synonymous with narratives based upon the conventional professional wisdom.

## Comparing and Contrasting the Approaches

Ernest R. House points out that any of the models can be appropriate or inappropriate depending on the circumstances of their application and the corresponding validity of the assumptions on which they are based. (23) Validity as the quality of being well-founded on fact, or established on sound principles, applicable to the case or circumstances under study and resulting in soundness, strength of argument, proof, and authority is a notion considerably expanded from the mere application of truth as a scientific, experimentally proved concept. Each of the eight approaches presented can make a claim of validity.

The models represent quantitative and qualitative processes for reaching judgments of merit and worth. None is more "scientific" than any other. The quantified approaches, harbor biases of their own; they are value ridden and the evaluators are not always aware of the biases.

The final four approaches described--art criticism, professional review, quasi-legal and case study--are qualitative. They base their claim to validity on an appeal to prolonged engagement and persistent review rather than quantified methods. (24) Observation is their primary method of data collection. Replication, a key criterion, is achieved by externalizing and explicating procedures so that events can be witnessed by several observers.

## Diversify

At this point in the evaluation of federally funded library programs, it is important that assessment is grounded in more than the simplistic idea that the application of a few input and output measures will lead to consistently valid judgments about program merit and worth. Nothing could be further from the truth. Contrary to the impression created by the literature, input and output measures are merely one part of a valid, reliable evaluation attuned to the individual library's context yet issuing data useful for national aggregation. The truth is that there are many valid approaches available to us which, if selected to fit the task and the situation, allow for diversity in evaluative design.

Too often we, as not only the creators, but also the evaluators of innovative library programs, misconceive our evaluative task and do an injustice to the programs we evaluate with an inadequate approach. Our continued insistence on using one hammer, labeled Input-Output, without proper attention to context, displays our naivete to the rest of the professional world.

## Notes and References

1. Thurow, Lester C. *The Zero Sum Society* (New York: Simon and Schuster, 1980) and Thurow, Lester C. *The Zero Sum Solution* (New York: Simon and Schuster, 1985) develop this theme.

2. Kosecoff, Jacqueline and Arlene Fink. *Evaluation Basics: A Practitioner's Manual* (Beverly Hills: Sage, 1982): 19-26.

3. DeProspo, Ernest R.; Ellen Altman; and Kenneth Beasley. *Performance Measures for Public Libraries* (Chicago: American Library Association, 1973).

4. Zweizig, Douglas and Eleanor Jo Rodger. *Output Measures for Public Libraries* (Chicago: American Library Association, 1982).

5. Van House, Nancy; Mary Jo Lynch; Charles R. McClure; Douglas L. Zweizig; and Eleanor Jo Rodger. *Output Measures for Public Libraries,* second edition (Chicago: American Library Association, 1987).

6. House, Ernest R. *Evaluating with Validity* (Beverly Hills: Sage, 1980): 18.

7. Patton, Michael Q. *Creative Evaluation,* second edition (Beverly Hills: Sage, 1987): 19.

8. House, Ernest R. "Assumptions Underlying Evaluation Models," *Educational Researcher* 7 (March 1978): 4-12 and House, Ernest R. *Evaluating with Validity:* 21-43; 225-247 provide comprehensive reviews of the approaches.

9. Rossi, Peter and Howard E. Freeman. *Evaluation: A Systematic Approach*, second edition (Beverly Hills: Sage, 1982): 18-59.

10. Hamburg, Morris; Richard Clelland; Michael Bommer; Leonard Ramist; and Ronald Whitfield. *Library Planning and Decision-Making Systems* (Cambridge: MIT Press, 1974).

11. Childers, Thomas and Nancy Van House. *The Public Library Effectiveness Study: Final Report to the U.S. Department of Education under Grant No. 008720290* (Philadelphia: Drexel University, College of Information Studies, 1989): 7-9.

12. O'Connor, Daniel. "Evaluating Public Libraries Using Standard Scores: The Library Quotient," *Library Research* 4 (Spring 1982): 51-70.

13. Lynch, Mary Jo. "FSCS and the Evaluation of LSCA." In: Betty J. Turock, Editor. *White Papers Series: Evaluating Federally Funded Public Library Programs* (Washington: U.S. Department of Education, Office of Educational Research and Improvement, Office of Library Programs, 1989). In press.

14. McClure, Charles R. "A View from the Trenches: Costing and Performance Measures for Academic Public Services," *College and Research Libraries* 47 (July 1986): 323-336.

15. Palmour, Vernon; Marcia Bellasai; and Nancy De Wath. *A Planning Process for Public Libraries* (Chicago: American Library Association, 1980).

16. McClure, Charles R.; Amy Owen; Douglas L. Zweizig; Mary Jo Lynch; and Nancy Van House. *Planning and Role Setting for Public Libraries* (Chicago: American Library Association, 1987).

17. Robbins-Carter, Jane and Douglas L. Zweizig. "Are We There Yet?" *American Libraries* 16 (October 1985): 624-627; 16 (November 1985): 724-727; 16 (December 1985): 780-784; 17 (January 1986): 32-36; 17 (February 1986): 108-112.

18. Dressel, Paul L. "Accreditation and Institutional Self-Study," *North Central Association Quarterly* 46 (Fall 1971): 277-287.

19. Estabrook, Leigh. "Accreditation of Public Libraries: An Outrageous Suggestion?" *Library Journal* 112 (September 15, 1987): 35-38.

20. Wolf, R. L. "The Use of Judicial Evaluation Methods in the Formulation of Educational Policy," *Educational Evaluation and Policy Analysis* 1 (May-June 1974): 19-28.

21. Lincoln, Yvonne S. and Egon G. Guba. *Naturalistic Inquiry* (Beverly Hills: Sage, 1985).

22. Miles, Mathew B. and A. Michael Huberman. *Qualitative Data Analysis: A Sourcebook of New Methods* (Beverly Hills: Sage, 1984).

23. Ernest R. House. *Evaluating with Validity.* 225.

24. Lincoln and Guba: 301-305.

# THE POTENTIAL ROLE OF PUBLIC LIBRARY ACCREDITATION FOR EVALUATING FEDERALLY FUNDED LIBRARY PROGRAMS

**Leigh Estabrook**

## Abstract

Accreditation, a voluntary self-regulatory process, has the potential to assist the federal evaluation process. It can provide information, first, about how a program will benefit from being carried out by the library requesting funds; and, second, about how the use of federal funds for a specific program will contribute to the improvement of the library. Current methods document programmatic contributions to efficiency and effectiveness, but do not assess the funded programs in relationship to the developmental status of the libraries which implemented them. A general overview of accreditation and of the work of the Ad Hoc Commission for the Accreditation of Public Libraries provides background information on the latest look at adapting the professional review process to the public library.

Accreditation has the potential to complement and extend evaluation tools currently used by the federal government and by the Public Library Association (PLA). The current approaches to evaluation, however conscientiously carried out, have one significant limitation: They focus attention on evaluation of specific proposals, without similar attention to the library in which the proposed programs will be carried out. Today's methods of evaluation can provide important information about the impact of federally funded programs on different user groups or the contribution of programs to administrative efficiency and effectiveness, but lack a complementary tool to assess the funded program in relationship to the institution of which it is a part. Before providing federal funding, it is important to know, first, how the program will benefit from being carried out by the specific library; and, second, how the program might contribute to the improvement of the institution as a whole. In the first instance, the following questions are posed: Is a public library capable of using the federal grant well? Are the staff and resources assigned appropriate to and the best available for the program. What is the likelihood of the program being continued at the end of the grant period? A well-written grant proposal will anticipate these question; but without professionally agreed upon criteria for assessment of the public library, it is difficult for a peer reviewer or a program officer to validate the library's answers.

More important, and even less easily assessed by current modes of evaluation, is the question of the extent to which a funded program contributes to the improvement of the library of which it is a part. Currently program evaluation focuses on the evaluation of outcomes. Did the program succeed or fail? Was the finished product what it was expected to be? It is equally important to ask about the impact of the particular program on the library's overall programs and mission. At the most basic level, did the grant support new programmatic efforts or were ongoing operating expenses reduced because grant funds were received? Did the new program generate new resources or did it require matching resources or support for overhead from the library's budget that might otherwise have been spent on higher priority services? How much did the grant funded program contribute to the overall mission and goals of the library? Are the priorities of the U.S. Department of Education congruent with those of the library or has the library shifted its priorities in order to be eligible for federal assistance? Federal and state evaluators cannot be faulted for not answering these narrower but crucial questions. As Crosson recently noted:

> Evaluation makes sense only where it applies to a process, exhibited in an object or institution, extended through time, and aimed at a definite term. (1)

## Peer Evaluation in Libraries

At present few methods exist for peer evaluation of libraries. The Public Library Development Project supported by PLA and the Council of State Library Agencies (COSLA) provides guidelines for library role and mission setting (2) and output measures (3) that would serve as excellent tools for library self-assessment, but there is no comparable provision for external validation that a library is in fact doing what it says it is doing or that a library is doing what is most appropriate for its community.

The State of Iowa has an accreditation process, but it is based on libraries meeting certain quantitative standards rather than on a process of self-assessment and peer review. The State of Vermont has moved significantly beyond this by developing a process entitled "Envisioning Excellence" that provides for external peer review of self-assessments submitted by public libraries. Although to date no libraries have been reviewed under this process, Vermont's libraries are working to complete the long-range plans and other components necessary for peer evaluation. (4) National public library accreditation could provide a broader framework for the evaluation of federally funded programs as well.

## Defining Accreditation

Accreditation is a voluntary self-regulatory process developed initially for the educational community. More recently it has been adopted by museums, prisons and police departments. Accreditation is designed to: 1) Recognize institutions or programs that have met or exceed standards, or established criteria for quality; and 2) Improve institutions or programs. Through the Council on Postsecondary Education, the federal government regulates accrediting bodies such as the Committee on Accreditation (COA) of ALA. Accreditation, itself, is not undertaken by the federal, state, or local government.

Accreditation standards are established by individual professions. In the more developed accrediting processes, standards are not numeric counts based on quantitative methodologies. Instead, they are designed to assess such qualities as the adequacy of resources, forms of management and facilities of the institution or program, and the responsiveness of the institution's programs to the communities it serves. These criteria provide the framework for an institutional self-study that is reviewed by the authorized nongovernmental accrediting body consisting of professional peers and representatives from the lay community.

The critical part of accreditation is not the final designation accredited or not accredited. Most valuable is the process of self-study that gives the institution the opportunity to examine the whole of its work. If the standards for accreditation are sound, staff involved in a self-study should have the opportunity to assess the strengths and weaknesses of their institution in a comprehensive way. Historical data, information from constituents served by the institution, and other externally generated information become part of the analysis. In both the self-study process and through peer review, the institution is forced to look at itself broadly and as others see it, and not only in the ways it might wish to be seen.

### Public Library Accreditation

To date, accreditation of public libraries exists only as a concept. It was rejected by the Board of Directors of the Public Library Association.

> The Public Library Association does not support the concept of accreditation developed by the Commission on Public Library Accreditation Ad Hoc Committee and therefore PLA will withdraw its official [sic] representative to the Commission and will communicate this decision to CAPL and to the PLA membership. (5)

That rejection stopped further dialogue on questions about accreditation that would help clarify the concept and its implementation. For example:

> ... Academic and school libraries participate in the accreditation process as subsets of larger institutions. How could the majority of small public libraries afford the time or expense of a self-study and site visit?
>
> ... If only large and wealthy public libraries can afford to go through the accreditation process and federal funding is, in some way, tied to accreditation, will we once again reward the wealthy?

As envisioned by the Ad Hoc Commission for the Accreditation of Public Libraries (CAPL), the objectives of public library accreditation would be consistent with the accreditation of other types of institutions in its overall goals: (1) To provide public assurance the programs and services of public libraries are of acceptable quality; (2) To assist public libraries in the improvement of their programs and services; and (3) To enhance public understanding of the contribution of public library programs and services to a community and encourage strong local, regional, and national support for those services.

Specific objectives for public library accreditation as articulated by CAPL include:

1. To provide an independent, autonomous agency to foster excellence in public libraries by developing, promoting, and applying standards and guidelines for assessing the effectiveness of public libraries in achieving their purposes.
2. To encourage public library improvement through continuous self-study and evaluation.
3. To require, as an integral part of the accrediting process, an institutional self-analysis that is analytical, interpretive, and evaluative, and an on-site review by a visiting team of peers.

4. To provide counsel and assistance to both developing and established public libraries.
5. To cooperate with various organizations representing public libraries for the purpose of maintaining and improving the best interest thereof.
6. To engage in such other activities necessary and proper for the accomplishment of these objectives consistent with the public interest and the interest of public librarianship. (6)

The specific criteria for accreditation that have been developed by CAPL emphasize qualitative achievement and are flexible enough to apply to public libraries of different sizes performing diverse roles in different types of communities. (7) For accreditation, libraries would provide evidence about their goals and objectives, planning and evaluation, governance and administration, collections and services, and financing and facilities. That evidence would be assessed within the context of the library's publicly stated long-term goals and specific objectives. A decision to grant accredited status would be based on an assessment of the library as a whole, not on the specific successes or failures of performance in one or another area. (8)

Public library accreditation does not substitute for, nor compete with, other modes of evaluation and, in fact, an effective accreditation process depends on the existence of well-considered professionally agreed upon standards and tools for evaluation. As Professor Peter Hiatt of the University of Washington's Graduate School of Library and Information Science has noted:

> If one .... accepts the basic assumptions underlying the work of PLA over the past decade, then accreditation becomes the logical next step: Define the role of the public library in the Untied States today (The Public Library Mission Statement and Its Imperatives for Service); create and utilize a planning process (the most current manifestation is *Planning and Role Setting for Public Libraries*); and evaluate the extent to which a library has utilized a, (not necessary the), planning process. (9)

Accreditation thus grows naturally out of the historical development of evaluation processes for libraries. It is recommended at this time because of the recent work of the Public Library Development Project. For example, in the proposed criteria for accreditation of public libraries, one criterion states, "The library's programs and other services are appropriate to and essential for the achievement of its objectives in meeting the needs of the community." The Public Library Development Project has made available the necessary tools for any size public library to examine the extent to which it meets this criterion.

## Public Library Accreditation and the Evaluation of Federally Funded Programs

How could public library accreditation--if implemented--assist in the evaluation of federally funded library programs? The major theme of this paper is that before providing support for a program a funding agency should know how the program will benefit from being carried out by the specific library and how the library will benefit from the program.

In the first instance, public library accreditation would ascertain whether a particular program can be carried out best by a particular library. That would provide one measure of quality assurance currently lacking. Federal and state agencies distributing grants could review proposals from libraries that hold accredited status with the knowledge that they have undergone external peer review and were found to use their resources effectively to achieve their objectives.

Anyone who has been responsible for proposal review will recognize the value of such an assurance from an external, neutral body. In that process--whether done by a federally sponsored reviewing panel or by a state advisory committee--proposal reviewers may tacitly draw on unsubstantiated information about the quality of individual libraries. As proposals from individual libraries are reviewed, such information plays into decision making. Hearsay, the experience of the reviewing party with the library or with its employees, or even knowledge about the community's attitude toward its public library can all affect the way in which a proposal is reviewed.

It is naive to assume that accreditation could eliminate all elements of subjectivity in the evaluation of proposals for federal funding, but it could help limit subjectivity in the important area of library quality. A library that has undergone the accreditation process would have externally validated evidence of the overall quality of its programs and services and would be able to provide broad evidence of what it, as an institution, might offer the funding agency.

This raises the sensitive issue of whether the federal government should encourage accreditation and then use the system as a mechanism for initial screening of applicants for funding. Accreditation of schools and colleges is currently a prerequisite for eligibility for federal funding to the institution. A similar requirement for public libraries might be considered once accreditation was operational, although it would probably take at least a decade to put in place a system which would be available to all public libraries seeking accreditation. The potential of using accreditation as an evaluation tool in the federal funding process is, however, one reason some members of the public library community have opposed the CAPL proposal. Achieving accreditation is seen as a complicated and expensive hurdle for those who seek federal support and, public libraries, aware of their weaknesses, worry that they might not be able to scale it.

The second objective of accreditation--to assist in the improvement of library programs and services--is one answer to those concerns. In the process of self-study and evaluation all libraries will identify areas in which improvement is desireable or needed. This information, too, could be effectively linked to the federal funding process. Although federal and state agencies set priorities for funding, those priorities do not always fit the priorities of need at the local level. For example, a library that is particularly advanced in using technology for interlibrary cooperation may not need Title III monies in the areas that are highest priority at this time, but there may have pressing local need in other areas. Rather than not granting funding to a library, thereby penalizing it for addressing national priorities with its own resources, it might be possible to target some federal funds to areas that need a significant improvement based on the accreditation assessment of that need at the local level. Of course, this would require amendment to current laws and/or regulations.

Federal funding targeted to areas needing improvement, as identified in the accreditation process, could strengthen the second goal of accreditation--library improvement--a goal explicitly shared with federal funding agencies. Programs could be evaluated in part on the extent to which they contribute to strengthening the library as an institution.

As described, accreditation could assist the evaluation of federally funded programs by providing evaluative information about the library during proposal review. In this case, it would most directly assist the evaluation of proposals, not the final evaluation of specific programs. Accreditation could also assist the broader goal of evaluation by helping identify areas within individual libraries toward which resources might most effectively be targeted and by providing a means to assess the effect of federally funded programs on the institution as a whole.

**Issues and Concerns**

The value of public library accreditation assumes that the process would focus on how the library succeeds or fails, not just which of the two it does. (10) Organizations that accredit schools and colleges are struggling over this issue. William Bennett, former Secretary of Education, argued that accreditation should be based more on educational outcomes than on the process of education. Outcomes accreditation has also been promoted by the North Central Accrediting Association which thought outcome measures should be used to supplement assessment of input measures and organizational process. (11,12)

To be truly useful to the evaluation of federally supported library programs, implementation of public library accreditation will have to address some of the problems that have been identified in existing accrediting processes, including:

1. The development in the public's mind of unrealistic expectations of the way in which the accrediting process can affect the institution.
2. The use of accreditation results as a political tool for internal and external organizational conflicts.
3. The degeneration of the accreditation process into descriptive conclusions that contribute only minimally to the goal of institutional improvement.

These problems are most likely to occur if the accreditation process becomes controlled by one or more interest groups, such as trustees, state libraries, or a professional association. For example, an accrediting body that operates through a professional association may face a choice between satisfying the association's membership and rigor in the accrediting process. The Commission on Public Library Accreditation began as an independent body precisely to avoid some of the politics that could compromise accreditation. As CAPL moves forward, however, it is imperative that it build an accrediting structure that involves the broadest community of interest, including both the professional and lay communities.

At present, the library community is divided in its beliefs about the value of accreditation. Strong support for the idea has been received from libraries of different sizes, even from some that are quite small. A number of trustees and community representatives, who are familiar with accreditation of other types of institutions, have also expressed their support. Librarians' concerns center on the cost--in time and money--for libraries to engage in a process with questionable return. But that is understandable, since prior to its completion, the self-study and accrediting review may have questionable value for institutional improvement.

Leadership from the federal government for public library accreditation can provide an important statement about accreditation's value. Linking accreditation to the evaluation of federally funded library programs would indicate to libraries the tangible benefits of participating in the accrediting process. At the same time the U.S. Department of Education could participate in the development of an important tool for the assurance of quality and the improvement of public library services in the United States.

## Notes and References

1. Crosson, Frederick J. "The Philosophy of Accreditation," *North Central Association Quarterly* 62 (Fall, 1987): 386-397.

2. McClure, Charles R. and others. *Planning and Role Setting for Public Libraries* (Chicago: American Library Association, 1987).

3. Van House, Nancy and others. *Output Measures for Public Libraries*, second edition (Chicago: American Library Association, 1987).

4. Communication from Marianne Cassell, Consultant, Vermont State Library, May 12, 1989.

5. PLA Board of Directors (1987-88) Document #51; Approved July 11, 1988.

6. Estabrook, Leigh. "Accreditation of Public Libraries: An Outrageous Suggestion?" *Library Journal* 112 (September 15, 1987): 35-38.

7. Criteria is used in preference to the more common term standards to avoid confusion with earlier input driven standards for public libraries that focused on quantitative measures of library quality.

8. Kimmel, Margaret M. and Leigh Estabrook. "Accrediting Public Libraries: An Update," *Library Journal* 113 (May 15, 1988): 54-55. The goals, objectives, and criteria of public library accreditation as proposed by CAPL are supplied here.

9. Letter from Peter Hiatt to Leigh Estabrook, August 24, 1987.

10. The value of this approach is strongly endorsed by Crosson in "The Philosophy of Accreditation."

11. Vaughn, John W. "An Evolutionary Revolution: Some Thought on Outcomes Accreditation," *North Central Association Quarterly* 60 (Winter, 1986): 386-397.

12. Wick, John W. "Including Assessment of Student Outcomes in the School Accreditation and Evaluation Process," *North Central Association Quarterly* 59 (Spring, 1985): 363-9.

# FSCS AND THE EVALUATION OF LSCA

**Mary Jo Lynch**

## Abstract

The Federal-State Cooperative System for Public Library Data (FSCS) was developed to coordinate the annual collection of public library statistics by state library agencies with the periodic reporting of national public library statistics by the National Center for Education Statistics (NCES). This paper describes how the system was initiated and how it can play an important role in the evaluation of federally funded programs, particularly the Library Services and Construction Act (LSCA), at the national, state, and local levels.

Over the last four years, I have been talking to people at state library agencies, at the National Center for Education Statistics (NCES), and to anyone else who would listen, about an idea that has been around for over a century. A system that would coordinate the annual collection of statistics from public libraries by state library agencies with the periodic reporting of national statistics on public libraries by NCES is very close to a reality now, thanks to the staff of many state library agencies, the staff of NCES, the National Commission on Libraries and Information Science (NCLIS), and even the Congress of the United States. (1) The system can play a role in the evaluation of federally funded library programs such as the Library Services and Construction Act (LSCA) at the local, state, and national levels.

My concept of evaluation includes all efforts, first, to measure or describe an entity and then compare that measurement or description to a previously established standard or an objective. My concern here is with evaluation of individual LSCA projects and of the program as a whole.

## The Ball Gets Rolling

The most recent idea for what eventually became the Federal-State Cooperative System for Public Library Data (FSCS) was born in the Fall of 1984 as I prepared "Analysis of Library Data Collection and Development of Plans for the Future." That project involved evaluation of the statistics collected by NCES from various types of libraries and any other agencies collecting library statistics. It also made recommendations for future action by NCES and prepared model questionnaires for NCES to use with each type of library. A summary of the final report appeared in the *Bowker Annual* for 1985. (2)

---

Acknowledgements: Thanks are due to Amy Owen, Utah State Library, and Jan Feye-Stukas, Minnesota State Library, who provided valuable suggestions for improving this paper.

The recommendations of that project covered academic, school and public libraries. A key component was a detailed analysis of the forms used by states to collect data from libraries. The three-volume appendix to the report contained charts that showed every item collected by any state and all states that collected the item. Two of the three volumes reported on public library statistics--a total of 377 pages. Clearly there was a wide variety in state data collection from public libraries in 1983-1984.

We found that all 50 state library agencies collected statistics annually from their public libraries. The general topics they gathered data on were similar, but the specific items were dissimilar and, therefore, the results were not comparable from state to state. The report recommended that NCES persuade the states to collect a limited set of key items in a standard way and report them to NCES so that national summaries could be created. Before the report was submitted, that idea was presented to the Chief Officers of State Library Agencies (COSLA) who supported it in principle.

Shortly after the November 1984 report was filed, the American Library Association sent a proposal to the U.S. Department of Education asking for funds to conduct a pilot project that would work with a small group of five to seven states to explore the feasibility of a system which would coordinate the annual collection of data from public libraries.

### Pilot Project

Two units of the U.S. Department of Education provided financial support for the project--Library Programs and NCES. In October 1985, an Advisory Committee was appointed, including Wes Doak, State Librarian of Oregon; Jan Feye-Stukas, Public Library Specialist, State Library of Minnesota; Amy Owen, State Librarian of Utah; Patricia Smith, State Librarian of Texas; and Barratt Wilkins, State Librarian of Florida. A letter went out immediately to all 50 chief officers of state library agencies inviting them to participant in the pilot. Twenty states volunteered--a much higher response than we had expected and a good omen for the future.

When the Advisory Committee met in November 1985 they decided to accept everyone who volunteered. By the time the project ended, there were 15 states officially participating: California, Colorado, Florida, Idaho, Indiana, Minnesota, New Hampshire, Ohio, Oklahoma, Oregon, Pennsylvania, South Carolina, Utah, Washington, and Wyoming.

At the November meeting the Advisory Committee also started to discuss items and definitions to be incorporated into existing state forms. States were free to gather whatever information they wanted as long as they collected a certain set of items using standard definitions and sent data on those items to NCES for incorporation into a national report. States could continue collecting data at whatever time of year was best for them. This would mean that aggregate reports would contain data collected at different times but sent to NCES at a standard time. In addition to making those decisions, the Advisory Committee spent a good bit of time revising the list of items and definitions that had been recommended in the report of the previous project. That list of definitions was to be revised two more times before the pilot project was finished.

In March of 1986 each state participating in the pilot project sent one or two representatives to a workshop in Chicago which covered such topics as forms design, how to get good data, and how to edit to remove errors. Workshop attendees also spent a good bit of time critiquing the list of items and definitions and recommending changes. After the workshop, the Advisory Committee and the Project Director prepared a revised list which was sent out to states in April for incorporation into the next cycle of data collection.

At the time of the March workshop only four states knew they could send data to NCES in machine-readable form--two on magnetic tape produced from a mainframe and two on floppies from a microcomputer. By the time the project was finished all were able to do so, albeit with different levels of skill and ease. This process was aided considerably by Gail McKenzie of the Indiana State Library and Dick Palmer of the Ohio State Library. Gail McKenzie gave us the first record layout showing what was done on a mainframe in Indiana. It became evident that several states planned to use Lotus 1-2-3. Dick Palmer of Ohio provided a Lotus format on paper and on diskette to any state that requested it. Twelve of the 15 pilot states sent 1986 data to NCES by the time the pilot project ended and the other three did so soon after that.

The pilot project was officially completed at the end of August 1987 when a final report was submitted which recommended that work begin immediately to expand to a 50-state system. By that time, Larry LaMoure had been appointed Library Statistics Coordinator at NCES. Both the final report on the pilot and Larry's appointment coincided with renewed interest in library statistics at NCES, prompted by a discussion of HR5 which eventually became Public Law 100-297 (the Hawkins-Stafford Elementary and Secondary School Improvement Act). This law specifically charged NCES to collect statistics about libraries. The need for a federal-state cooperative system for public library data is specifically mentioned in the law.

In February of 1988, a Memorandum of Understanding (MOU) was signed between NCES and NCLIS authorizing NCLIS to coordinate the work of a Task Force in initiating the expansion. That Task Force met monthly from March through September 1988 and developed a plan for FSCS. One key Task Force decision was to reduce the pilot project's list of 81 data elements to 41. The other basic ideas remained the same, however. (3)

## Going Into Action

Implementation of the plan began with a workshop for state data coordinators in Annapolis on December 5-8, 1988. Forty-eight states and the District of Columbia sent representatives to learn what was required of participants and what help was available from NCES and NCLIS.

The original Task Force continues to provide general guidance for FSCS. Closer to daily operations is the Implementation Committee of staff from state library agencies, NCLIS, and NCES. A Technical Committee of state library staff also advises the project on matters related to computers and data processing.

As of May 1, 1989, 40 states were expected to submit their most recent data to NCES by July of 1989. There will undoubtedly be some problems and much work remains to increase 40 states to the total 50, but the FSCS database will exist very soon.

## The Current and Future Composition of the Database

To date, basic indicators of public library development make up the database. They are derived from 41 core elements in ten broad categories, including general information, site identifiers, employers, income and expenditures, capital outlay, library collections, public service hours, service per typical week, circulation, and interlibrary loans.

The FSCS 1989 Action Plan, published in May of 1989, recommends that after several years of using the data elements and definitions, the participants should work to incorporate indicators in nine additional areas:

1. Population of legal service area: Develop definitions and standard methodologies for determining the population served by a reporting library. These definitions should accommodate public library systems, cooperatives, and federations that serve a portion of an individual library's service area population.
2. Contractual services: Develop definitions that standardize the reporting of contractually supplied services such as bookmobiles and rotating film collections. This will eliminate duplicate reporting of services, materials, and expenditures.
3. Central and branch libraries: Consider elimination of the distinctions between central and branch libraries in the data file, which may not be necessary after the development of the universe file.
4. Capital and operating expenses: Evaluate the variations among states of definitions of capital expenses, and if warranted, develop new definitions to ensure uniformity of data.
5. Physical facility space: Consider collecting and reporting this information.
6. Registered borrowers: Experience may demonstrate that a greater number of public library reporting institutions can accurately supply this information as more libraries adopt computerized circulation systems. If the agencies are able to supply this information accurately, begin collecting and reporting it.
7. Automated services: Consider making surveys of automated support services in libraries and automated database services offered by public libraries.
8. Titles/volumes: Study, and adopt if feasible, the collection of data identifying the number of different titles as well as the number of volumes.
9. Telecommunications: As more public libraries use information technology and digital communications systems consider collecting such data. This would require the development of standard definitions for telephone, FAX, and telecommunications capabilities.

## FSCS and Evaluation at the National Level

LSCA exists because it was possible to convince the Congress that public library service is a good thing for the American people, that it increased their ability to govern themselves and enhanced their lives in many ways. (4) Ideally an evaluation of LSCA would examine whether self-government and the quality of life have improved because of LSCA, but that kind of study would be extremely complex. Although methodologies exist, using them requires great skill and a long period of time, both of which translate into dollars. What is possible, and much less costly, is to produce statistics that describe the size and shape of the public library enterprise in the United States on a periodic basis. Methodologies here are much less complex, because we are dealing with facts not with the perceptions and attitudes of human beings. With descriptive statistics, federal officials and the library community can know basic facts about public libraries and can monitor changes over a period of years in factors like hours open, items circulated, and reference questions answered. Such information provides a solid framework within which evaluation of the aspects of library programs funded by LSCA can be conducted.

At this moment, the most recent national data on basic aspects of public library service is seven years old. (5) Once FSCS is in operation, current national data will be available and it will be updated regularly. At any time then, evaluators at the national level can know the basic facts about the libraries to which LSCA funds are being applied and can watch those facts change. It would be simplistic to claim that LSCA alone was responsible for any change that might be observed, especially since this legislation has never provided more than a very small percentage of funding for public library service nationally. However, it can be assumed that LSCA is one of the casual factors, and changes could have implications for the future implementation of LSCA. The bottom line is that obtaining basic descriptive statistics about the public libraries enterprise will facilitate an intelligent use of federal funds. FSCS will produce those figures on a timely basis so that large gaps between national data reports, like the current one of 7 years, will not exist in the future.

Another way in which the development of FSCS should help in the evaluation of LSCA is less direct. On the road to an operational FSCS, we encountered several potholes of confusion regarding some of the basic concepts used in talking about public libraries. To begin with, what is a public library? LSCA defines it as, "A library that serves free of charge all residents of a community, district or region and receives its financial support in whole or in part from public funds," but that definition doesn't answer such questions as these: Is the public library the building on the corner, or the system of buildings of which that place on the corner is a part? Is a state library a public library? Does the answer to that question change if the state library provides bookmobile services to rural areas otherwise without public library service? If a community taxes itself for public library service and has a public library board but that board chooses to provide local service by contracting with other communities, does that community have a public library? After encountering numerous issues like this, the Task Force working on FSCS developed a taxonomy of entities providing public library service. That taxonomy is being used now by state library agencies to identify the public library entities in their states in order to contribute to the public library universe file that is part of FSCS. Once the universe file exists, those administering LSCA will have a much clearer picture of the nature of the institutions they are funding.

Population served was another concept that needed clarification. Traditionally, public libraries are described and compared in terms of population served, but that way of thinking may be an anachronism today because of the development of systems; the practice of communities contracting with other communities for some or all library services; statewide library cards; transportation and work patterns that make it convenient for people to use libraries other than the one for which they are taxed. The Task Force worked out a definition of population served for use in FSCS, but Task Force members are aware that it does not solve all the problems.

One problem, unsolvable to date, is how state library agencies determine what proportion of the state's population is served by public libraries. Discussion with others on the Task Force led me to hunt for statistics. I found them in the *American Library Directory* (ALD), but what I found only increased the dilemma. The introductory pages for each state in ALD give some basic statistics which are supposedly supplied by state library agencies. A table was constructed to display three of those statistics: Population, population served by public libraries, and unserved. Forty-nine states provided some or all of these items. Of those 49, 25 states admitted to having some unserved population.

Eighteen gave a specific figure for the unserved. Seven others did not give a figure, but there was a difference between the figures for population and for population served by public libraries. One of the states that gave a figure for unserved, added a note that they are "served by mail by the State Library with In-Watts telephone access."

Of the 24 states that did not admit to any unserved population, 20 gave the same figure for both population and for population served by public libraries. Three states left the population served by public libraries column blank. Since the unserved column is also blank in those cases, these states were counted as not admitting to any unserved. One state did not give a figure, but said, "all by county and local libraries and books by mail."

Looking at those results raises several questions. Is it possible that every person in the 24 states is served by public libraries? How do the 25 states, admitting to unserved population, determine what it means to be unserved? Is books-by-mail really public library service? If the Congress asks what percentage of the population is served by public libraries, what will we tell them? These questions all point to the key issue: How do we define what we mean by access to library service? Is there a difference between some access and adequate access? How do states deal with this issue in the long-range plans required by LSCA?

Related to population served is the issue of how to account for system services. One of the major uses of population served is to group libraries for purpose of comparison. We have traditionally assumed that libraries serving the same size population ought to be similar in terms of budget, collection size, staff size, level of circulation, and number of reference questions answered. But is it legitimate to compare two libraries serving the same population when one stands alone and the other receives many services from a system? If system services were taken into account, would it be possible to differentiate various levels of system service and various sources of system support? Since a primary focus of many LSCA projects has been the development of library systems, this is an important distinction. FSCS has not solved these dilemmas, but they are being faced and results should benefit LSCA evaluators by providing a better understanding of public library service and organization.

## State Level Evaluation Improvements

At the state level, FSCS can contribute toward both long-range program development and subsequent evaluation activities related to LSCA. First of all, participating in FSCS will improve the quality of data about public libraries collected by the state library agency. That particular benefit was discovered during the pilot project. When a number of states were very late in meeting the deadline for sending their data to NCES, the representatives of pilot project states attending the ALA Conference in San Francisco confessed they had examined the data more carefully and found anomalies that had to be corrected by checking back with the local libraries. Because these states were now part of a national project, they were unwilling to accept data that had been "good enough" earlier for state purposes. We can expect this to happen in many states as they join FSCS. Thus state agencies will have better data for all purposes related to local public libraries, including planning for uses of LSCA, evaluation of proposals for LSCA funding, and reports on results of LSCA projects.

States will also have comparable data for other states. Although the major purpose of FSCS is to gather data for national reporting, an important secondary purpose is to enable individual states to exchange data sets. A particular state may not really care what the national picture is for a particular statistic but may be very interested in what is happening in a state with similar demographic and geographic characteristics. Such sharing is facilitated when both states are FSCS participants. Already a group of five Western states have formed an organization, the Western Interstate Library Data Cooperative, WILDCAT for short, for cooperative analysis of data. The cooperative is based at the Colorado State Library under the leadership of Keith Lance, Director of the Library Research Program.

Another benefit of FSCS for states is the increased capacity for using microcomputers to manage and analyze data. That might have happened anyway, as almost everyone dealing with data is getting a PC and learning how to use spreadsheets or database management software. However, the pilot project accelerated the trend in participating states by providing a reason to begin using micros and help in doing so. Now that NCES is fully committed to FSCS, there is even more help available.

Having the equipment and skill to send data to NCES for incorporation into the national statistics should have numerous spillover effects at the state level that will help in evaluation activities related to LSCA. Not only will basic descriptive statistics be easy to manipulate when appropriate but the staff will have information and skills that can be transferred to separate data collection efforts needed for LSCA evaluation.

## Local Level Benefits for Evaluation

Participation in FSCS will also have benefits at the local library level. When thinking about this point, it is important to realize that a wide variety of institutions are included when we speak of public libraries. According to the 1981-82 NCES public library statistics, 58 percent of the 8,597 public libraries had less than $50,000 to spend in that year; 55 percent of them had a full-time professional staff of less than two; and 53 percent of them had collections of less than 20,000 volumes. The overlap among libraries in those categories is probably high. At the other extreme, according to the same statistics, were the 381 (4.4 percent) with operating expenditures of more than $1 million, the 37 (.4 percent) with staff greater than 100, and the 55 (.6 percent) with collections of over one million. (6) When we speak of local libraries we are talking about both extremes and all those in between. The ways in which FSCS will help these institutions prepare LSCA grant proposals and evaluate results of LSCA projects will, of course, differ. For the very small libraries with low budgets, staff, and collections, participating in FSCS will mean that staff will get special training from the state library agency with regard to basic descriptive statistics. Experience with the pilot project leads me to suspect that some small libraries will produce accurate statistics for the first time because of FSCS. Although local staff may see this as a mixed blessing, attention to basic statistics should ultimately prove beneficial whether or not LSCA funds are involved.

Larger libraries are expected to have staff with expertise to deal with statistics, but that it not always true and help from state library staff may be welcome. Comparable data from large libraries in other states may also be welcome. Since there are many fewer large libraries than small libraries, a large library may have to go to another state to make comparisons. FSCS will ensure that comparable statistics are maintained in different states.

## LSCA Evaluation Enhanced

LSCA has been in existence for a long time and FSCS is just beginning. If the two had developed in tandem, evaluation of the long-term impact of LSCA on the nation's public libraries would be a lot easier today. For the future, however, FSCS will be there, providing basic descriptive statistics useful at the local, state, and national levels in evaluating individual projects and the program as a whole. FSCS cannot stand alone as a tool for evaluating LSCA, but it will provide a quantitative infrastructure of consistent basic data upon which specialized and/or qualitative evaluation can be built.

## Notes and References

1. Lynch, Mary Jo. "An Impossible Dream Come True," *Public Libraries* 28 (Winter, 1988): 170-171.

2. Lynch, Mary Jo. "Analysis of Library Data Collection and Development of Plans for the Future: A Project Summary." In: *Bowker Annual of Library and Book Trade Information* (New York, R.R. Bowker, 1985): 405-411.

3. *An Action Plan for a Federal-State Cooperative System for Public Library Data (FSCS).* Developed by the Task Force on a Federal-State Cooperative System for Public Library Data with the support of NCLIS and NCES. (Washington, DC: National Commission on Libraries and Information Science, April 1989)

4. Holley, Edward G. and Robert F. Schremser, *The Library Services and Construction Act. An Historical Overview From the Viewpoint of Major Participants* (Greenwich, CT: JAI Press, 1983): 16.

5. *Statistics of Public Libraries, 1981-82.* Data gathered by the National Center for Education Statistics, U.S. Department of Education. (Chicago: Public Library Association, 1985)

6. *Statistics of Public Libraries.* For expenditures, see Table 18: 104; for staff, see Table 11: 94; for collection, see Table 21: 108.

# THE NATIONAL DIFFUSION NETWORK: ITS POTENTIAL FOR LIBRARIES

**Ellen Altman and Philip M. Clark**

## Abstract

The National Diffusion Network (NDN), within the U.S. Department of Education's Office of Educational Research and Improvement (OERI), disseminates information about successful educational programs and provides teacher training so that innovations can be replicated and utilized by other schools. Programs seeking to join the NDN undergo a review and validation by the Program Effectiveness Panel (PEP). Although library programs are eligible to participate, to date none have been submitted to PEP scrutiny. Programs funded by LSCA and denoted exemplary are described in *Check This Out*, but none contain evaluation designs that would meet PEP standards. Rather than relying exclusively on classical experiments for proof of effectiveness, PEP criteria have been broadened to accept other methods, depending on the types of claims about effectiveness made for the programs. Four types of claims are discussed along with the evaluation methods acceptable for each. Recommendations are made for improving the evaluation designs in LSCA funded projects based on the NDN model.

Although proposals submitted for funding under the Library Services and Construction Act (LSCA) have for many years required applicants to complete a section on how their proposed projects would be evaluated, evaluation remains the weakest part of most proposals, both submitted and funded. According to Guy Garrison and Galen Rike, "Evaluation is something that is promised in an application but rarely delivered in a final project report." (1)

The validity of Garrison and Rike's assertion is substantiated by *Check This Out*. (2) This is a 1987 publication intended to disseminate information about exemplary library and media center programs selected by a panel of reviewers. The book is composed of excerpts from the reports of 62 exemplary programs, most of which had received federal funds.

Of the 20 programs in the group supported at least partially by LSCA, only five included a section labeled Evaluation. Of these, three gave no information about the number of people reached nor any evidence about the impact of the services provided. One said there had been no time to evaluate. The fifth offered evidence that had no relationship to the objectives of the funded program.

Nine of the LSCA-funded programs did not specifically mention the word evaluation, but they did include a brief section that might be labeled results. These sections were named: "How Is the Program Doing?" "Estimating the Effect of the Changes," "Program Effectiveness," "Impact of Program," or "Meeting Needs." Most of these gave circulation or rate of growth figures for certain elements of their programs. Some evidence offered in support of the goodness of the programs was only tangential to the nature or purpose of the programs. For example, one school district which received funds to consolidate and operate its audiovisual services used as evidence the scores achieved by students on the library-related sections of the ARS Achievement Tests without explaining how the two were connected.

Another six reports lacked an evaluation section or any indication of results. However, three of these did include some statistical information on library usage. One which gave statistics on its contacts, including phone calls to a recorded dial-a-message service, failed to note that its unit costs were over $9 per contact.

Too many of the reports offered vague statements to support the positive impacts of their programs, such as, "Community response indicates that the program provided important benefits to human service organizations and to the general public." Several others relied on only one letter of testimony from a satisfied user.

Few of the LSCA programs included in *Check This Out* even described what they hoped to accomplish beyond being able to offer a program or a special collection. The unmentioned assumptions are that their programs are inherently good and that their availability is inherently valuable. The program initiators and the reviewers have failed to recognize the difference between intentions and outcomes or effort and effect. One of the original expectations for *Check This Out* was that evaluative criteria and comparative data might emerge from currently applied methodology, methods, and measures. To date that expectation has not been realized.

Since the evidence presented for efficacy of so many of these programs is weak, the question arises: What can the U.S. Department of Education, which funded these programs, initiate to help improve the indicators of quality and rigor for evaluating library-related proposals? Perhaps it can use, as a model a highly regarded mechanism already existing within the Department --the National Diffusion Network. Although *Check This Out* was funded by the Recognition Division of the U.S. Department of Education, "to promote linkages between the National Diffusion Network (NDN)....and the library community," the Introduction carries the disclaimer that the evaluation data presented by these libraries would not qualify their programs for approval by the NDN. (3)

## How the NDN Works

The U.S. Department (then Office) of Education created the National Diffusion Network in 1974 to disseminate information about successful educational programs developed with federal funding and to provide teacher training so that programs could be replicated and utilized by other schools. These represent only a tiny fraction of the number of proposals funded annually by the Department. To be judged successful a program must be more than innovative; it must prove its validity and significance.

The sources of innovation are the people who have created, field tested, and measured the educational improvements resulting from their programs. These demonstrator/developers, as they are called, are willing to provide training, materials and assistance to schools who wish to apply their methods. Each state has a facilitator who acts as the linking point between demonstrator/developers and the schools. The facilitators make local schools aware of innovations and help select and implement specific NDN programs. (4) The individual schools bear the costs of adopting the programs they select.

The value of NDN programs to the adopting schools and their faculties is proven by the statistics on program adoption for 1988 alone. That year, 26,088 schools paid to adopt 153 approved programs which were used for 3,271,803 students. In addition 72,035 teachers and 5,948 school administrators were trained in utilizing programs. (5) This training usually involves workshops, actual use of the program for at least one year, and a demonstration of competence.

102

The Recognition Division of the U.S. Department of Education directs and administers funds for dissemination of NDN programs, including arranging for an annual publication--*Educational Programs That Work*--which lists all currently approved projects. (6) However, the Division's most important task is to work with the Program Effectiveness Panel (PEP), formerly known as the Joint Dissemination Review Panel (JDRP). This review group judges whether the claims made by programs are sufficiently valid and worthy for inclusion in the National Diffusion Network. The 60-member Program Effectiveness Panel has included representatives from professional associations, school systems, and higher education along with staff from the U.S. Department of Education who have experience in educational research, teaching effectiveness, and evaluation. Programs developed without federal funding are now also eligible for consideration. Information about PEP prepared by the Office of Educational Research and Improvement stresses that the Program Effectiveness Panel does not evaluate programs [per se].

> Rather PEP validates the evaluations done by others. An approval by PEP means that the evidence presented before the panel warrants the evaluation claims that a program achieves specific results. A disapproval does not necessarily reflect poorly on the program; disapproval usually reflects poorly on the evaluation evidence. A disapproval means that the evidence presented to the panel does not warrant the evaluation's conclusions. (7)

Demonstrator/developers who wish to apply for inclusion in NDN submit a written statement of not more than 15 pages which follows the format specified in the criteria and guidelines handbook. According to Dr. Stanley Pogrow, Associate Professor of Educational Administration at the University of Arizona, "You have to show that what you have developed is better than the conventional approaches."

Dr. Pogrow's program to increase the higher order thinking skills of disadvantaged students was approved by the PEP panel in 1988. He said that although the applicants must demonstrate that educational gains are statistically significant, significance alone is not enough; "You also have to prove effect sizes." Dr. Pogrow's initial application was turned down two years ago, in part, because he had test results from only one site. "You have to have a lot of data, not only about effectiveness, but also about disowning alternative hypotheses. You have to show that the improvements are not due to some quirk." Furthermore, applicants must explain the educational importance of the results in terms of fulfilling needs and in comparison to other similar programs. Dr. Pogrow said, "The panel asked very good questions. They picked the application apart pretty well. Essentially, the panel tries to find reasons to argue that the data do not meet the claim." (8)

The application statement is first reviewed by the appropriate U.S. Department of Education program office. The reviewer from that office decides whether the submission should be rejected, revised, or reviewed. The review is done by mailing the submission to six members of the panel. According to regulations published in the *Federal Register* notification 34 CFR 786, August 14, 1987, the panel must award points in accordance with the following criteria:

Results (0-50 points). The program must clearly state the need being met and the intended purpose of the program. The results of the program must show an explicit connection between the observed changes and the need.

Evaluation Design (0-40 points). The evaluation design must be appropriate for the program. It must demonstrate that a clear connection exists between the evidence presented for the desired educational outcome, which is directly attributable to the program, and account for rival hypotheses that might explain effects.

Replication (0-10 points). The program should be adaptable to other schools with the strong likelihood of achieving similar results (9). Also, the time, money and other resources required must be realistic in terms of the results. The total points given by each panel member are averaged for the final score. A project's admission to the NDN requires at least an average score of 40 for the results section and an overall total of 70 points. The chairperson reviews the panel's written comments about denied applications whose total scores fall between 50 and 69 to see whether more evidence or clarification of certain points would justify another review by an in-person panel. The project developer is invited to attend this meeting to present additional evidence for approval.

Dr. Pogrow stressed that projects passing the PEP review do not necessarily receive any funding. That decision is made in another open competition and constrained by previous funding commitments. However, he likens PEP approval to getting the Good Housekeeping Seal, "It gives your project credibility and widespread interest." (10)

## Can Library Programs Pass PEP Scrutiny?

In fact, there are no administrative barriers keeping library programs, whether funded by LSCA or not, from approval by the NDN. The barrier appears to be lack of knowledge about NDN both on the part of the librarian program developers and the libraries that might be interested in adopting programs. A convenience sample taken by the authors of this paper indicated that none of the librarians we queried had ever heard of the National Diffusion Network! That barrier can be removed by disseminating information about the NDN at conferences and in the professional literature.

A more serious obstacle is getting library-related programs, especially those funded by LSCA, through the rigorous PEP approval process. Sarah Jane Roberts of the RMC Corporation, wrote a thoughtful position paper, "Evaluating Library Programs for the NDN." She believes that, "Library program evaluation stands now where educational program evaluation stood 15 years ago." (11)

A prima facie case can be made that school-related programs have several advantages in evaluating their programs which public libraries, the primary focus of LSCA funding, lack. These include a captive student audience, the ability to test freely, and better standards of comparison based on validated educational norms. However, as Roberts points out, educational evaluation has benefited substantially because of the continuing refinement of both methodology and standards of evaluation developed over the years. She believes, "Library programs need and deserve the guidance of the PEP in improving their evaluation practices and applying the concept of educational significance to their outcomes." (12)

## Claims and Criteria Set Up by PEP

The PEP model of program evaluation is rigorous. Until the issuance of the latest criteria and guidelines as reflected in *Making the Case* (13), the substantiation of a claim of program effectiveness relied heavily on classical evaluation designs, typically the experimental-control group, pre-test/post-test design. This type of design is difficult to achieve in many of the programmatic environments of libraries. But more critically, it is the availability of trained evaluators as part of the program staff that permits such designs to be attempted. Lynch has shown that the presence of an independent evaluator affiliated with a research firm is a prime determinant of approval. (14)

The latest version of the criteria and guidelines for PEP panelists extends the permissible evaluation designs. No longer is the rigorous classical design the only design. Now different criteria are used depending on the claims that are made as to how effective the program has been. This is important for library program designers because it explicitly recognizes the limitations of classical design for many of the programs offered by libraries.

A claim of program effectiveness is a statement that a result was achieved and that the result is educationally significant. Four types of claims are now recognized by the PEP administrators; along with criteria and guidelines, they are detailed in *Making the Case*. (15)

Presentation of hypothetical library claims can illustrate possible linkages with the NDN.

**Claim Type 1.** Claims about academic achievement--that the student recipient had a significant change in knowledge or skills as a result of the program:

Typically, such situations might occur in library bibliographic instruction programs. A claim of this type would be that students who participated in bibliographic instruction attained significantly higher scores on a standard test than did nonparticipants. To substantiate such a claim, it would be necessary to match participants and nonparticipants to show that they started with like levels of knowledge--ascertained by a test--and that the situation was controlled so as to limit any measurable effects to the program alone. In other words, a classical experimental design would be needed. In addition, the significantly higher scores would be in the range of a one-third or higher change in the standard deviation of the pre- and post-test scores. Literacy and database searching programs could also meet this claim.

**Claim Type 2.** Claims that improvements were made in teachers'--or in this case librarians'--attitudes and behaviors:

Here the claim is not that knowledge and skills increased but rather that intermediate effects on attitudes and/or behaviors changed significantly for the positive. For example, reference librarians who participated in a program on nonjudgmental thinking were rated significantly higher on approachability by minority clients. The critical results desired here phrased in noninstructional terms, are increases in the amount of assistance given, increases in the amount of time devoted to assistance, changes in methods of assisting patrons, and positive changes in librarians' attitudes toward patrons. PEP panelists are instructed to look beyond questionnaires to structured interviews, structured observations and unobtrusive measures as substantiation of Type 2 claims. When questionnaires are used, nonresponse bias must be taken into account, a factor that is rarely recognized in library survey research. The primary test of the design will be made on how well it proves that it was the program and not other factors that caused the difference in the participants' attitudes and behaviors. This criteria is constantly mentioned in the PEP guidelines.

**Claim Type 3.** Claims that the program resulted in improvements in students'--or in this case patrons'--attitudes and behaviors:

As with a Type 2 claim, the intention here is not a direct change in knowledge but rather a change in attitudes or behaviors in a specific, targeted subgroup not a whole population. For example, prison inmates view themselves as self-learners through exposure to the program, or program participants increase their visits to the library or borrow more materials, or attitudes toward reading-for-fun are significantly more positive than before and in a comparison group.

Again, it must be shown that the program itself made the difference. An important piece of evidence might be longitudinal data showing persistence of change in the participants. While somewhat less rigorous, the evidence must still be of a quality rarely shown in typical library research much less library demonstration programs.

**Claim Type 4.** Claims about improvements in instructional practices and procedures:

Instead of noting changes in individuals (librarians, teachers, students, patrons), this area focuses on institutional factors. Programs that reduce costs or increase efficiency, improve service to particular client groups, promote cooperation, or provide new service are included. The following is one example of a Type 4 claim:

> Increase in use of resources and facilities: One year after conversion of a neglected branch library into a homework facility staffed by teacher-librarians and stocked with young-adult level materials, monthly figures for library visits quadrupled, the number of library cards issued doubled, and circulation was three times larger. (16)

As stated in the guidelines, "Claim Type 4 is appropriate when the project meets the following conditions: It is aimed at the immediate effect of producing changes in the school, system, or institution, and/or changes in a general population or service area; it consists of a coherent set of procedures than can be transferred to similar institutions; and it postulates that the outcomes will contribute to student achievement some time in the future." (17)

The criteria and guidelines acknowledge that the major problem with this type of claim and its substantiation is the comparison standard. The comparison standard is less rigorous than in the classical experimental design where a control group is present. In this case, the comparison can take place in one of two ways. First, the competitive practice situation offers comparison of the program to like programs in like situations. It is anticipated that most programs are not really unique but rather are modifications and enhancements of standard practice. Therefore, there needs to be proof that this program achieved substantially better results, or cost significantly less, or drew overwhelmingly greater audiences than does standard practice in libraries similar to this one. For example, you might have to show that circulation gains over a year were greater than in other comparable, high-performing libraries.

The second situation is the unique practice situation where no comparison exists. That is, the program has never been tried before. Therefore, program success will probably be judged on how much things have changed from the way there were before the program. The degree to which such change is judged to be significant will be determined by what sounds impressive in the minds of PEP panelists. And the $9 cost per telephone request as mentioned earlier does not sound impressive in a positive sense.

As Roberts and the authors of *Making the Case*, Ralph and Dwyer, point out, a body of comparison data will emerge as projects come before the panel. Even rejected projects will contribute to the definition of a standard of comparison. Needless to say, the library community must submit programs to build this critical body of knowledge.

## Adding Library Programs to the NDN

As of April 1989, no federally funded library program had been submitted to the PEP for acceptance by the NDN. The new guidelines should make it somewhat easier for many projects to qualify, but only if careful consideration of evaluation design criteria are made an element of the programs and their planning. The profession should not exclude experimental designs from its thinking, even though they are most difficult to achieve. But realistically, most projects might be judged positively if they adhere to the less stringent standards of Type 2, 3, and 4 claims.

The designs must address the matter of appropriate comparisons either with other programs or with prior conditions. Project developers must conclusively show that it was the effects of their programs and not extraneous effects that produced the claimed results. This is done by detailing what happens when the program is not present. Instruments such as questionnaires, interviews, and tests, must be shown to be valid and reliable. And, in the end, it must be shown that results were significant and substantial.

Many of these requirements are technical and draw upon a knowledge of research methodology that is usually not required in our professional education. This lack of training in research design does not restrict librarians from relying on authoritative testing organizations, such as Consumers Union, for information to pass on to their patrons. But this type of testing rigor is not demanded for the adoption of library programs. It seems that claims of effectiveness like the "how I did it good" articles in library literature rely primarily on intuitive evidence. Librarianship deserves better.

## Moving Forward

LSCA is basically a program administered by the 50 state library agencies in accordance with each state's plan for library development submitted to the U.S. Department of Education. The Library Development Office within each state creates its own instructions and forms for LSCA applications. According to federal regulations, an advisory council in each state, composed of librarians and interested citizens, selects the proposals recommended for funding. The Office of Library Programs at the U.S. Department of Education helps interpret the regulations and monitors compliance. However, the states generally make no coordinated effort to disseminate information about innovative projects. News about some programs gets communicated at state or national conferences or in the professional literature. But there is no mechanism to validate the success of programs or to help other libraries replicate them.

NDN operates as a confederation of local, state, and federal partners working to improve schools and learning. A confederation of agencies similar to that of NDN already exists for libraries--local libraries, the state library agency, and the Office of Library Programs in the U.S. Department of Education. The office in each state library agency which handles LSCA grants could function in the same manner as the state facilitators do for NDN. Over half of the facilitators work in state departments of education. The people in charge of state library development could be given the responsibility of identifying those small number of proposals having the potential to meet PEP criteria. This would allow good evaluation plans to be worked out, perhaps with the expertise of a consultant, before the programs are actually implemented. Since this would be done for only one or two programs per year for each state, the routine procedures for awarding LSCA funds would remain the same. The library development officers could publicize exemplary programs funded in their own states, keep abreast of exemplary programs in other states, and connect demonstrators/developers with libraries wishing to adopt programs.

Implementing such change would require a substantial and continuing training program along with an effort to re-orient attitudes among some state library personnel. David Shavit, who has made an extensive study of federal programs and the state library, lists a number of reasons why improving LSCA evaluation would be a formidable task. These include lack of agreement about the definition and conceptual framework of evaluation, questions about the worth of evaluation in terms of the time and work required, fear about challenges to the worth of programs, criticisms that training in evaluation is too theoretical, and continuing turnover of staff. (18)

As a beginning toward recognizing programs that really work well, potentially strong programs with a high probability of acceptance by PEP and NDN could be identified by a national panel of practitioners and educators. This group could be headed by a staff member from the Office of Library Programs who is an expert in evaluation and, preferably, a member of the American Evaluation Association. This expert would assist with evaluation designs and implementation or recommend the help of other consultants before funded proposals are actually begun. The expert would review applications for the PEP review. Completed projects that have been judged exemplary should be repeated elsewhere but with appropriate evaluation designs included from the beginning.

The profession has enough experts to prepare proper evaluation designs for the relatively few numbers that would be identified as high payoff programs. They are library educators, systems analysts, and program officers. The larger problem of educating those with great program ideas in the techniques of evaluation is a longer and much more difficult task. But illustrating that it can be done through the mechanism of adoption by NDN should be a high priority of the profession.

Libraries have produced valuable and significant results with programs they have planned and implemented. These programs should be verified and validated by a recognized body of experts. The end result can only be better, cheaper and more usable program ideas for the profession and our users.

**Notes and References**

1. Garrison, Guy and Galen E. Rike. *An Assessment of Selected Aspects of the Administration of LSCA Projects by the Indiana State Library* (Muncie, Indiana: Ball State University, Department of Library Science, 1979): 23.

2. *Check This Out: Library Program Models.* Prepared for the National Diffusion Network under contract to RMC Corporation, contract 300-84-0251. (Washington, D. C.: Department of Education, Office of Educational Research and

Improvement, 1987).

3. *Check This Out*: 1.

4. LoGuidice, Tom. "The National Diffusion Network: A Report on a Resource for America's Elementary and Secondary Schools," *Clearing House* 59 (September 1985): 33-34.

5. U.S. Department of Education, Office of Library Programs, Figures supplied by Betty J. Turock in consultation with Lee Wickline, Director of the National Diffusion Network, April 1989.

6. *Educational Programs That Work* (Longmont, CO: Sopris West, annual).

7. Ralph, John and M. Christine Dwyer. *Making the Case: Evidence of Program Effectiveness in Schools and Classrooms - Criteria and Guidelines for the U.S. Department of Education's Program Effectiveness Panel* (Washington, D.C.: U.S. Department of Education, Office of Educational Research and Improvement, November 1988) DRAFT: 2-3.

8. Interview with Dr. Stanley Pogrow, University of Arizona, April 12, 1989.

9. Ralph and Dwyer. *Making the Case.*

10. Pogrow, Interview.

11. Roberts, Sarah Jane. "Evaluating Library Programs for the NDN: A Position Paper." Unpublished, n.d.: 1. Ms. Roberts was affiliated with the RMC Corp. at the time *Check This Out* was prepared.

12. Roberts: 1.

13. Ralph and Dwyer. *Making the Case.*

14. Lynch, Kathleen Bodisch. "Practices in Educational Program Evaluation, 1980-1983." Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C.: April 20-24, 1987).

15. Ralph and Dwyer. *Making the Case.*

16. Ralph and Dwyer: 36.

17. Ralph and Dwyer: 34.

18. Shavit, David. *Federal Aid and State Library Agencies: Federal Policy Implementation* (Westport, CT : Greenwood Press, 1985): 125.

## ABOUT THE CONTRIBUTORS

**ELLEN ALTMAN and PHILIP CLARK,** both of whom hold the Ph.D. from Rutgers University, have collaborated previously on *A Data Gathering and Instructional Manual for Performance Measures for Public Libraries* (Chicago: Celadon, 1976) with Ernest R. DeProspo and Ellen O'Connor Clark. Altman is Professor and former Director of the Graduate Library School, University of Arizona. Her major publications include *Performance Measures for Public Libraries* with Ernest R. DeProspo and Kenneth Beasley (Chicago: Public Library Association, 1973), and *Local Public Library Administration,* second edition (Chicago: American Library Association, 1980). Clark is currently Associate Professor of Library and Information Science at St. John's University. Formerly Director of the Montclair (NJ) Public Library, he is the author of *Microcomputer Spreadsheet Models for Libraries* (Chicago: American Library Association, 1985) and a regular columnist for *The Bottom Line: A Financial Magazine for Librarians* (New York: Neal-Schuman, quarterly).

**MARY K. CHELTON,** Library Programs Coordinator, Montgomery County Department of Public Libraries, Rockville, Maryland, earned an M.L.S. from Rutgers University and an M.P.H. from the University of Alabama's School of Public Health. She has served as President of the American Library Association's Young Adult Services Division and Chair of the Public Library Association's Service to Children Committee. With Dorothy Broderick, she founded the journal, *Voice of Youth Advocates* (VOYA). Chelton was a consultant to the national study on resources and services for young adults in public libraries, recently completed by the U.S. Department of Education's National Center for Education Statistics (NCES).

**GEORGE D'ELIA,** Associate Professor in the Information and Decision Sciences Department at the University of Minnesota's Carlson School of Management, received the M.L.S. and Ph.D. from Rutgers University and the M.S. in Statistics from Syracuse University. He has served as Director of the Graduate Library School at the University of Minnesota, taught at the School of Information Studies at Syracuse University, and worked as a systems analyst and Head of the Circulation Department at the University of Colorado Library. His areas of expertise include planning and evaluation of library systems and services, survey research methods, and statistics. D'Elia's research addresses the measurement of library user behavior, library use, and the performance and evaluation of library services. He won the Library Research Round Table Competition for the best research paper of the year in 1980, 1982, and 1984. A member of the editorial board of the *Journal of Academic Librarianship* since 1982, he has served as a consultant to public libraries, government agencies, and profit and nonprofit corporations. He has conducted more than 30 workshops in the United States and Canada on library user surveys, library performance evaluation, and statistical analysis of library data.

**LEIGH ESTABROOK,** Dean and Professor at the Graduate School of Library and Information Science, University of Illinois, Urbana-Champaign, holds the Ph.D. in sociology from Boston University and the M.S. in Library Science from Simmons College. Among Estabrook's recent consulting clients are Harvard University, Graduate

School of Design and Murphy/Jahn Architects. She is Chair, Ad Hoc Commission for the Accreditation of Public Libraries. Her most recent publication is a study of the growth of the profession which appeared in *College and Research Libraries* (May 1989). Estabrook's current research, funded by a grant from the University of Illinois Research Board and the Association for Library and Information Science Education (ALISE), includes a study of the effect of technological change on the nature of library work and the structure of the library workforce. She is also directing an interdisciplinary team of scholars in an investigation of Scholarly Communication and Information Transfer partially funded by a Beckman Research Award--given to projects of special distinction or special promise by the Research Board of the University of Illinois.

**MARY JO LYNCH** completed her Ph.D. at Rutgers University. Since 1978 she has held the position of Director of the Office of Research at the American Library Association. She has worked as a reference librarian in academic libraries, has taught in three library education programs, and is a frequent contributor to the professional literature. She served as Project Coordinator for the work that led to publication of *A Planning Process for Public Libraries* (Chicago: American Library Association, 1980), and on the Steering Committee that guided the development of the first edition of *Output Measures for Public Libraries* (Chicago: American Library Association, 1982).

**CHARLES R. McCLURE**, President of Information Management Consultant Services, Inc. and Professor at the School of Information Studies, Syracuse University, earned his Ph.D. at Rutgers University. Principal investigator on the Public Library Development Project, he is a frequent management consultant to public, academic, corporate, and state libraries. He is author, co-author or editor of 19 monographs and over 100 reports and articles, including *Planning for Library Services* (New York: Haworth Press, 1982); *Strategies for Library Administration* (Littleton, CO: Libraries Unlimited, 1982); *Research for Decision Making* (Chicago: American Library Association, 1984); *Planning and Role Setting for Public Libraries* (Chicago: American Library Association, 1987); and *Output Measures for Public Libraries*, second edition (Chicago: American Library Association, 1987). Currently, McClure is completing work, with Nancy Van House and Beth Weil, on *Output Measures for Academic and Research Libraries*.

**BETTY J. TUROCK**, Chair of the Department of Library and Information Science at Rutgers University, received her Ph.D. there. She has done additional graduate study in clinical psychology at the University of Pennsylvania. Prior to entering library education, Turock was Assistant Director, then Director of the Montclair (NJ) Public Library and Assistant Director of the Rochester and Monroe County (NY) Library System. She is the author of more than 40 publications on the public library, including the books *Serving the Older Adult: A Guide to Library Programs and Information Services* (New York: R.R. Bowker, 1982); and *Financial Planning for Libraries* (New York: Neal-Schuman, 1989). The founding editor of *The Bottom Line: A Financial Magazine for Librarians* (New York: Neal-Schuman, quarterly), she is currently completing *Evaluation for Library Managers* for Scarecrow Press.

**NANCY VAN HOUSE** holds the Ph.D. from the School of Library and Information Studies, University of California, Berkeley, where she is currently Associate Professor. She has also served as Senior Research Associate at King Research, Inc.

The author of numerous books and articles, including *Public Library User Fees* (Westport, CT: Greenwood Press, 1983), she co-authored (under the name DeWath) *A Planning Process for Public Libraries* (Chicago: American Library Association, 1980); *Planning and Role Setting for Public Libraries* (Chicago: American Library Association, 1987); and *Output Measures for Public Libraries*, second edition (Chicago: American Library Association, 1987). Her major areas of research and consulting are library planning and evaluation, the economics of library services, and the library labor market.

**DOUGLAS L. ZWEIZIG**, Professor, School of Library and Information Studies, University of Wisconsin-Madison, teaches in the areas of research methods, library management, communications, and public libraries. He completed his Ph.D. at Syracuse University. Zweizig is co-author of *Output Measures for Public Libraries*, both the first and second editions (Chicago: American Library Association, 1982 and 1987); and *Planning and Role Setting for Public Libraries* (Chicago: American Library Association, 1987). He has directed projects that adapted output measures for use with children's services and literacy programs.